LAWRENCE
LIVERMORE
NATIONAL
LABORATORY

LLNL-JRNL-448115

An Improved Convergence Analysis of the Smoothed Aggregation Algebraic Multigrid

M. Brezina, P. Vanek, P. S. Vassilevski

August 9, 2010

Numerical Linear Algebra with Applications

Disclaimer

AN IMPROVED CONVERGENCE ANALYSIS OF THE SMOOTHED AGGREGATION ALGEBRAIC MULTIGRID

MARIAN BREZINA, PETR VANĚK, AND PANAYOT S. VASSILEVSKI

Abstract. We present an improved analysis of the smoothed aggregation (SA) algebraic multigrid method (AMG) extending the original proof in [SA] and its modification in [Va08]. The new result imposes fewer restrictions on the aggregates that makes it easier to verify in practice. Also, we extend a result in [Van] that allows us to use aggressive coarsening at all levels due to the special properties of the polynomial smoother, that we use and analyze, and thus provide a multilevel convergence estimate with bounds independent of the coarsening ratio.

1. Introduction

The smoothed aggregation (or SA) algebraic multigrid (or AMG) has become one of the methods of choice for solving large sparse linear systems for equations that typically arise from discretizing elliptic partial differential equations. This paper focuses on the multilevel analysis of the method by improving the original result presented in [SA] and its modification in [Va08]. The convergence result that we prove in the present paper imposes only one restriction on the aggregates formed at every level; namely, to have a diameter bounded above by a specific number that depends only on the polynomial degree used in the construction of the prolongation matrix. For general aggregation strategies, the bound also generally depends on the shape of the resulting aggregates (which is hidden in the constant of the Poincaré inequality that we assume for our model setting). For the specific general aggregation strategy that we outline in a later section, that constant can be shown to be uniformly bounded. The final result is qualitatively similar to the original one, i.e., the relative condition number of the V-cycle SA AMG operator with respect to the original matrix is of order ℓ^3 where ℓ is the number of levels involved in the cycle.

An additional contribution of the paper is that the presented analysis allows for aggressive coarsening, and the proved convergence bound is independent of the coarsening ratio, provided that a special polynomial smoother is employed as multigrid relaxation. The smoother we use is a symmetrization of the one used previously in [VBT, Van]. We present a new analysis of this smoother that makes it possible to incorporate aggressive coarsening at all levels.

We note that multilevel methods featuring aggressive coarsening typically suffer deterioration of convergence unless a more powerful smoothing is employed. A very common

Date: January 18, 2010, today is July 8, 2010.

1991 *Mathematics Subject Classification.* 65F10, 65N20, 65N30.

This work was performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344.

choice in such cases is to resort to using overlapping Schwarz smoothers. The polynomial smoother employed here offers an alternative that is expected to be cheaper than the use of overlapping Schwarz smoothers. This is the case, since we show that the polynomial degree ν needed to compensate for large coarsening factor H/h, where $h \ll H$, are the mesh sizes of two consecutive levels, is sufficient to satisfy $\nu = \mathcal{O}(H/h)$. This was first recognized, for a two–level method in [KrVa] and [VBT] and used in a V/W-cycle (for aggressive coarsening between the first and second level only) in [Van]. For problems posed on a d–dimensional domain of unit size, to implement one action of such polynomial smoother, we need order $\frac{H}{h}N$ operations, where $N = \mathcal{O}(h^{-d})$ is the number of unknowns at the given fine level. The storage requirement is $\mathcal{O}(N)$ which is optimal (it is already accounted for since we keep the matrices anyway). The cost of a typical Schwarz smoother is readily estimated to be of order H^{-d} (number of subdomains) times the cost to factorize, store and solve with the individual blocks. For the best direct factorization method such as nested dissection, we have the following well-known (cf., e.g., [TD06]) cost estimates for a finite element matrix on a $m \times m$ ($\times m$) mesh in 2D ($3D$):

- "cost of factorization:"
$$\begin{array}{rl} \mathcal{O}\left((m^d)^{\frac{3}{2}}\right), & d = 2, \\ \mathcal{O}\left((m^d)^2\right), & d = 3. \end{array}$$
- "storage requirement:"
$$\begin{array}{rl} \mathcal{O}\left(m^d \log m\right), & d = 2, \\ \mathcal{O}\left((m^d)^{\frac{4}{3}}\right), & d = 3. \end{array}$$
- "cost for triangular factors solve:" (same as the number of nonzeros of the triangular factors)
$$\begin{array}{rl} \mathcal{O}\left(m^d \log m\right), & d = 2, \\ \mathcal{O}\left((m^d)^{\frac{4}{3}}\right), & d = 3. \end{array}$$

A straightforward calculation shows that the Schwarz method requires more than $\mathcal{O}(N)$ setup cost to compute the triangular factors of the Schwarz blocks. More specifically, in $3D$ we have, with $m = H/h$, $\mathcal{O}(H^{-d})$ (the number of Schwarz blocks) times $\left((H/h)^d\right)^2$ (the cost of nested dissection to factor one block) which gives a total of $\mathcal{O}(\nu^3 N)$ operations (in 3D). In $2D$, we have $\mathcal{O}(\nu N)$ factorization cost. Also, the storage requirement in both, $2D$ and $3D$, is suboptimal. More specifically, to store the non-zero entries of the (lower) triangular factors of the Schwarz blocks it requires $\mathcal{O}(H^{-d})$ times $\mathcal{O}\left((H/h)^d \log H/h\right)$ which gives $\mathcal{O}\left(N \log \nu\right)$ storage in $2D$, and $\mathcal{O}(H^{-d})$ times $\mathcal{O}\left(\left((H/h)^d\right)^{\frac{4}{3}}\right)$, that is, $\mathcal{O}(\nu N)$ storage in $3D$. Likewise, to solve linear systems with the triangular factors, we need $\mathcal{O}(N \log \nu)$ in $2D$, which is better than $\mathcal{O}(\nu N)$ needed for the polynomial smoother, whereas in $3D$ it is $\mathcal{O}(N\nu) = \mathcal{O}\left(H^{-d}\left((H/h)^d\right)^{\frac{4}{3}}\right)$ ($d = 3$), which is of the same order as the cost of the polynomial smoother. Another point to stress here is that generalizing the existing (geometry based) analysis of Schwarz smoothers is difficult to carry out in a multilevel setting for algebraic multigrid coarse spaces due to their complex geometric properties.

Finally, to fill–in the gap of the previous convergence results, namely, to have the complexity of one V-cycle under control, we suggest one (geometrically based) aggregation procedure, that after the first level of coarsening generates matrix graphs with regular structure (similar to a finite element matrices on uniformly refined meshes) which then gives us a straightforward tool to estimate the cost of the V-cycle and the setup cost of generating the AMG hierarchy (prolongation and coarse-level matrices).

The remainder of the paper is structured as follows. In Section 2, we formulate the SA method by introducing the relevant notation and tools. Section 3 reviews some properties of an optimal polynomial that are constructed on the basis of the standard Chebyshev polynomial, and formulate some technical results needed in the main analysis.

Section 4 contains the main assumptions for our analysis, namely a standard weak approximation property formulated in a matrix-vector form. The analysis itself is found in Section 5. Section 6, contains the definition and analysis of the polynomial smoother that ensures convergence result independent of the coarsening ratio in cases of very aggressive coarsening. We complete the paper with Section 7 that contains one geometric strategy for generating aggregates that would produce coarse matrix graphs with regular structure and illustrate it with some numerical tests combined with the polynomial smoother introduced in Section 6. Finally, as an Appendix, we provide some technical estimates used in the analysis.

2. Preliminaries

To be specific, in the present paper we consider sparse $n \times n$ symmetric positive definite (or s.p.d.) matrices A that come from a finite element discretization of the model second order elliptic partial differential equation (or PDE)

$$-\nabla \cdot a(x)\nabla u = f, \tag{2.1}$$

posed on a domain Ω, a polygon (in 2D) or polytope (in 3D), with Dirichlet or combination of Neumann and Dirichlet boundary conditions. The given coefficient $a = a(\mathbf{x})$ satisfies

$$0 < a_1 \leq a(\mathbf{x}) \leq a_2 \text{ on } \Omega,$$

and the given right hand side, f, is in $L_2(\Omega)$.

The domain is triangulated by a mesh, or set of elements, $\mathcal{T}_h$ and with $\mathcal{N}_h$ we denote the respective set of (nodes) vertices of elements in $\mathcal{T}_h$. In the simplest case of piecewise linear polynomial functions used to define the respective finite element space V_h, the degrees of freedom is the set $\mathcal{N}_h$. Discretizing the above PDE using the Galerkin method with this finite element space and its usual Lagrangian (nodal) basis, we obtain the system of linear algebraic equations,

$$A\mathbf{u} = \mathbf{f}, \tag{2.2}$$

the solution of which is our main interest. The $n \times n$ matrix A is symmetric positive definite (s.p.d.) and sparse, i.e., having overall $\mathcal{O}(n)$ nonzero entries (or more precisely $\mathcal{O}(1)$ nonzero entries per row). Its sparsity naturally defines a graph, which in our simple case of lowest order elements coincides with the mesh $\mathcal{T}_h$. More specifically, for any nonzero entry a_{ij} of A, we have an edge of the matrix graph that connects its vertices i and j. Nonzero entries are possible only if two nodes $\mathbf{x}_i$ and $\mathbf{x}_j$ in $\mathcal{N}_h$ are vertices of a

common element $\tau \in \mathcal{T}_h$. (To be precise, we consider as nonzero entries some possibly numerical zero entries.)

We will use ℓ_2-vector norm, i.e., $\|\mathbf{v}\| = \sqrt{\mathbf{v}^T\mathbf{v}}$ and for a given s.p.d. matrix X, the norm $\|\mathbf{v}\|_X = \sqrt{\mathbf{v}^T X\mathbf{v}}$.

Our interest is in analyzing the smoothed aggregation (SA) algebraic multigrid as a nearly optimal (in convergence properties) method for solving problem (2.2).

Given the finest-level matrix, A, the SA method bases its coarsening on a non-overlapping partition of the set of fine-level nodes, $\mathcal{N}_0 = \mathcal{N}_h$, into sets, called aggregates, $\{\mathcal{A}_i\}_{i=1}^{n_c}$. At this point, we do not specify any particular algorithm for generating aggregates. As with other algebraic multigrid methods, SA must ensure that error components not sufficiently attenuated by relaxation (referred to as "*algebraically smooth*" components) must be removed by the coarse-grid correction. For the simple model problem and the relaxation schemes considered here, such algebraically smooth error components are locally constant (discretizations of Poisson operator have stencils with zero row sum in the interior of the domain). Thus, in this simplest case, SA uses the constant function 1, or its coefficient vector $\mathbf{1}$, which restricted to each aggregate defines a block–diagonal matrix $I_c = \text{diag}\ (c_i\mathbf{1}_i)$; $\mathbf{1}_i = \mathbf{1}|_{\mathcal{A}_i}$ and $c_i = \frac{1}{\|\mathbf{1}_i\|}$. I_c is referred to as a tentative prolongator. In the case considered here, it amounts to a piecewise constant interpolation, which, when used as a multigrid transfer operator, as is well-known, leads to highly non-scalable convergence of the resulting multigrid V-cycle. That is why the tentative prolongator is not used by SA directly. Instead, the actual prolongator (or interpolation matrix) P is obtained as a product of a certain matrix polynomial with smoothing properties, times I_c. The coarse level matrix, A_c, is defined variationally via the Galerkin relation $A_c = P^T AP$. We also define the coarse vector $\mathbf{1}_c = (\frac{1}{c_i}) \in \mathbb{R}^{n_c}$. We have, by construction, that I_c has orthogonal columns of unit size, i.e., $I_c^T I_c = I$ and that $\mathbf{1} = I_c\mathbf{1}_c$. We identify the coarse degrees of freedom (row and column indices of A_c) with the indices of the set of aggregates $\{\mathcal{A}_i\}_{i=1}^{n_c}$. They also define the vertices $\mathcal{N}_c$ of the matrix graph associated with A_c.

We let $A_0 = A$, $\mathbf{1}_0 = \mathbf{1}$, $n_0 = n$, $\mathcal{N}_0 = \mathcal{N}_h$, and similarly, $A_1 = A_c$, $\mathbf{1}_1 = \mathbf{1}_c$, $n_1 = n_c$, $\mathcal{N}_1 = \mathcal{N}_c$, $I_1^0 = I_c$, and $\mathcal{A}_i^0 = \mathcal{A}_i$ for $i = 1, \ldots, n_c$.

The coarsening proceeds by recursively creating aggregates at every consecutive level, $k \geq 1$, of the multigrid hierarchy (larger k stands for coarser level). For this to be possible, we only need to have available, at the current level, the set of aggregates $\{\mathcal{A}_i^{k-1}\}_{i=1}^{n_k}$ (generated at the preceding level) or equivalently the set of nodes $\mathcal{N}_k$ (defined as the indices of the preceding level aggregates $\mathcal{A}_i^{k-1}$), the matrix A_k, and the vector $\mathbf{1}_k$. At the current level k, we form a new set of aggregates $\{\mathcal{A}_i^k\}_{i=1}^{n_{k+1}}$ by grouping together nodes $\mathcal{N}_k$ of level k. The current-level aggregates are sets of connected nodes that provide a non–overlapping partition of $\mathcal{N}_k$. The connectivity is defined via the graph (sparsity pattern) of the matrix A_k.

The current level vector $\mathbf{1}_k$ is restricted to each aggregate $\mathcal{A}_i^k$ denoted by $\mathbf{1}_{k,\,i}$ and the tentative prolongator I_{k+1}^k is defined as the block diagonal matrix $\text{diag}\ \left(c_i\ \mathbf{1}_{k,i}|_{\mathcal{A}_i^k}\right)_{i=1}^{n_{k+1}}$. The constant $c_i = \frac{1}{\|\mathbf{1}_{k,i}\|}$ is a scaling factor used to make the columns of I_{k+1}^k on unit size, and serves to define the next $(k+1)$-level vector as $\mathbf{1}_{k+1} = (\frac{1}{c_i}) \in \mathbb{R}^{n_{k+1}}$.

Next, we define the actual prolongator (or interpolation matrix) P_{k+1}^k as a product of a matrix polynomial times I_{k+1}^k. That is,

$$P_{k+1}^k = S_k I_{k+1}^k, \tag{2.3}$$

where

$$S_k = \varphi_\nu(A_k), \tag{2.4}$$

for a polynomial $\varphi_\nu(t)$ of given degree $\nu \geq 1$ that is normalized at the origin, i.e., $\varphi_\nu(0) = 1$. Our analysis, exploits certain optimal, Chebyshev–based, polynomial, to be introduced in Section 3.

The coarse-level matrix is defined by the variational (Galerkin) formula,

$$A_{k+1} = (P_{k+1}^k)^T A_k P_{k+1}^k. \tag{2.5}$$

With A_{k+1}, the set of aggregates $\{\mathcal{A}^k\}_{i=1}^{n_{k+1}}$ constructed and hence the set of coarse nodes, $\mathcal{N}_{k+1}$, identified, and also the vector $\mathbf{1}_{k+1}$ defined, the recursion can proceed in the same manner on the next coarse level.

The above procedure can be illustrated more explicitly, as follows. At level k we construct

$$I_{k+1}^k = \begin{bmatrix} c_1\mathbf{1}_{k,1} & 0 & 0 & \dots & 0 \\ 0 & c_2\mathbf{1}_{k,2} & 0 & \dots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \dots & 0 & c_{n_{k+1}-1}\mathbf{1}_{k,n_{k+1}-1} & 0 \\ 0 & 0 & \dots & 0 & c_{n_{k+1}}\mathbf{1}_{k,n_{k+1}} \end{bmatrix} \begin{matrix} \} \, \mathcal{A}_1^k \\ \} \, \mathcal{A}_2^k \\ \} \, \vdots \\ \} \, \mathcal{A}_{n_{k+1}-1}^k \\ \} \, \mathcal{A}_{n_{k+1}}^k \end{matrix},$$

$$\mathbf{1}_{k,i} = \mathbf{1}_k|_{\mathcal{A}_i^{k+1}}, \quad c_i = \left(\mathbf{1}_{k,i}^T \mathbf{1}_{k,i}\right)^{-1/2}, \quad i = 1, \dots, n_{k+1}, \text{ and } \mathbf{1}_{k+1} = \begin{bmatrix} c_1^{-1} \\ c_2^{-1} \\ \cdot \\ \cdot \\ \cdot \\ c_{n_{k+1}}^{-1} \end{bmatrix}.$$

The construction ensures that $\mathbf{1}_k = I_{k+1}^k \mathbf{1}_{k+1}$ and $(I_{k+1}^k)^T I_{k+1}^k = I$.

Next, we introduce the composite tentative prolongator

$$I_k = I_1^0 \ \dots \ I_k^{k-1} \text{ for } k \geq 1, \quad I_0 = I. \tag{2.6}$$

It is clear that $\mathbf{1} = \mathbf{1}_0 = I_k \mathbf{1}_k$ for every $k \geq 0$. By construction, we also have $I_k^T I_k = I$ and hence $I_k I_k^T : \mathbb{R}^n \mapsto \mathbb{R}^n$ is an ℓ_2–orthogonal projection. Introducing the restriction matrix $Q_k = I_k^T : \mathbb{R}^n \mapsto \mathbb{R}^{n_k}$, then $I_k Q_k$ is the ℓ_2–projection of our further interest.

Similarly, we can introduce the composite aggregates $\widetilde{\mathcal{A}}_i^k$. They are defined as the following sets of the finest-level degrees of freedom (nodes)

$$\widetilde{\mathcal{A}}_i^k = \text{support } \left(I_k \chi_{\mathcal{A}_i^k}\right),$$

where χ_T stands for the characteristic function of the set T viewed as a vector of zeros and ones. In other words, the composite aggregate $\widetilde{\mathcal{A}}_i^k$ is the support of the vectors in the range of the composite tentative prolongator with domain restricted to the space of vectors supported in the aggregate $\mathcal{A}_i^k$.

Alternatively, we may define the composite aggregates in the following recursive way:

$$\widetilde{\mathcal{A}}_i^k = \cup\{\widetilde{\mathcal{A}}_j^{k-1} : \ j \in \mathcal{A}_i^k\}.$$

To define a standard V-cycle multigrid, in addition to the hierarchy of matrices $\{A_k\}$ and prolongators $\{P_{k+1}^k\}$, we need the smoothing matrices M_k. We assume that M_k provide iterative methods with error propagation operator $I - M_k^{-1}A_k$, such that $M_k^T + M_k - A_k$ is coercive in terms of A_k and also we assume the scaling, uniformly in $k \geq 0$, for a constant $\alpha > 0$,

$$\alpha \ \mathbf{v}_k^T A_k \left(M_k^T + M_k - A_k\right)^{-1} A_k \mathbf{v}_k \leq \mathbf{v}_k^T A_k \mathbf{v}_k \text{ for all } \mathbf{v}_k \in \mathbb{R}^{n_k}. \tag{2.7}$$

Furthermore, we assume that M_k satisfies one of the bounds, either M_k is spectrally equivalent to Richardson, i.e., $M_k = b_k I$ for an explicitly available norm bound $b_k \geq \|A_k\|$ (then (2.7) trivially holds), or in addition to (2.7),

$$\mathbf{v}_k^T \overline{M}_k \mathbf{v}_k \leq \beta \ \left(b_{k+1}\|\mathbf{v}_k\|^2 + \mathbf{v}_k^T A_k \mathbf{v}_k\right), \tag{2.8}$$

where b_{k+1} is an upper bound of $\|A_{k+1}\|$ that will be specified later and β is a positive constant.

Here and throughout the paper

$$\overline{M}_k = M_k \left(M_k^T + M_k - A_k\right)^{-1} M_k^T, \tag{2.9}$$

stands for the symmetrized smoother. It can be defined implicitly from the relation

$$I - \overline{M}_k^{-1} A_k = (I - M_k^{-T} A_k)(I - M_k^{-1} A_k).$$

The estimates (2.7) and (2.8) are verified for the polynomial smoother that we use to handle the case of large coarsening factor.

Based on a given choice of P_{k+1}^k, M_k (that is A_k-convergent), for $0 \leq k \leq \ell - 1$, and A_k obtained variationally from A_{k-1} for $1 \leq k \leq \ell$, starting with $B_\ell = A_\ell$, for $k = \ell - 1, \ \ldots, \ 1, \ 0$, we recursively define a V–cycle preconditioner (a s.p.d. matrix) B_k in the following standard way:

$$I - B_k^{-1} A_k = (I - M_k^{-T} A_k) \left(I - P_{k+1}^k B_{k+1}^{-1} (P_{k+1}^k)^T A_k\right) (I - M_k^{-1} A_k).$$

Letting $B = B_0$, we are concerned in what follows with the (upper) bound K_* in the estimate

$$\mathbf{v}^T A \mathbf{v} \leq \mathbf{v}^T B \mathbf{v} \leq K_\star \ \mathbf{v}^T A \mathbf{v}. \tag{2.10}$$

3. An optimal Chebyshev-like polynomial

Here, we review a polynomial with certain optimal properties (see, [BD96], [Va08], see also [BHMV] and [Sha94]). Its construction is based on the classical Chebyshev polynomials.

Consider the Chebyshev polynomials $T_k(t)$ defined by recursion as follows, $T_0 = 1$, $T_1(t) = t$ and for $k \geq 1$, $T_{k+1}(t) = 2tT_k(t) - T_{k-1}(t)$. Letting $t = \cos\alpha \in [-1, 1]$, we have the explicit representation $T_k(t) = \cos k\alpha$, which is seen from the trigonometric identity $\cos(k+1)\alpha + \cos(k-1)\alpha = 2\cos\alpha\cos k\alpha$.

We now summarize some properties of T_k (the proofs are straightforward and can be found in [Va08]).

Proposition 3.1. *For a given $b > 0$, the function defined for $t \in [0,\ b]$*

$$\varphi_\nu(t) = (-1)^\nu \frac{1}{2\nu+1}\ \frac{\sqrt{b}}{\sqrt{t}}\ T_{2\nu+1}\left(\frac{\sqrt{t}}{\sqrt{b}}\right), \tag{3.1}$$

is a polynomial of degree ν such that $\varphi_\nu(0) = 1$, that is, $\varphi_\nu(t) = 1 - tq_{\nu-1}(t)$ for some polynomial $q_{\nu-1}(t)$ of degree $\nu - 1$.

Proposition 3.2. *The polynomial φ_ν defined in* (3.1) *has the following optimality property:*

$$\min_{p_\nu:\ p_\nu(0)=1}\ \max_{t\in[0,\ b]}\ |\sqrt{t}\ p_\nu(t)| = \max_{t\in[0,\ b]}\ |\sqrt{t}\ \varphi_\nu(t)| = \frac{\sqrt{b}}{2\nu+1}. \tag{3.2}$$

Also, $\varphi_\nu(0) = 1$ and

$$\max_{t\in[0,\ b]}\ |\varphi_\nu(t)| = 1. \tag{3.3}$$

Here are some particular cases of the polynomials φ_ν.
Using the definition of the Chebyshev polynomials, $T_0 = 1$, $T_1 = t$, $T_{k+1} = 2tT_k - T_{k-1}$, for $k \geq 1$, we get $T_2 = 2t^2 - 1$ and hence

$$T_3(t) = 4t^3 - 3t.$$

Thus,

$$\varphi_1(t) = -\frac{1}{3}\sqrt{b}(4\frac{t}{b^{\frac{3}{2}}} - \frac{3}{\sqrt{b}}) = 1 - \frac{4}{3}\ \frac{t}{b}.$$

This in particular shows that

$$\sup_{t\in(0,\ b]}\ \frac{|1-\varphi_1(t)|}{\sqrt{t}} = \frac{4}{3}\ \frac{1}{\sqrt{b}}.$$

The next polynomial is based on $T_5 = 2tT_4 - T_3 = 2t(2tT_3 - T_2) - T_3 = (4t^2 - 1)(4t^3 - 3t) - 4t^3 + 2t = 16t^5 - 20t^3 + 5t$. Therefore,

$$\varphi_2(t) = \frac{1}{5}\sqrt{\frac{b}{t}}\left(16\sqrt{t}t^2\frac{1}{b^{\frac{5}{2}}} - 20\sqrt{t}t\frac{1}{b^{\frac{3}{2}}} + 5\sqrt{t}\frac{1}{\sqrt{b}}\right).$$

This shows,

$$\varphi_2(t) = \frac{16}{5}\frac{t^2}{b^2} - 4\frac{t}{b} + 1.$$

We also have,

$$\sup_{t\in(0,\ b]}\ \frac{1-\varphi_2(t)}{\sqrt{t}} = \frac{4}{\sqrt{b}}\ \sup_{x\in(0,1]}\ (x - \frac{4}{5}x^3) = \frac{4}{3}\sqrt{\frac{5}{3}}\frac{1}{\sqrt{b}}.$$

In general, it is clear that the following result holds.

Proposition 3.3. *There is a constant C_ν independent of b such that the following estimate holds,*

$$\sup_{t\in(0,\ b]}\ \frac{|1-\varphi_\nu(t)|}{\sqrt{t}} \leq C_\nu\frac{1}{b^{\frac{1}{2}}}. \tag{3.4}$$

Proof. We have, $1-\varphi_\nu(t) = tq_{\nu-1}(t)$, that is, $\frac{1-\varphi_\nu}{\sqrt{t}} = \sqrt{t}\; q_{\nu-1}(t)$ and therefore the quotient in question is bounded for $t \in (0, b]$. More specifically, the following dependence on b is seen:

$$\sup_{t\in(0,\,b]} \frac{|1-\varphi_\nu(t)|}{\sqrt{t}} = \frac{1}{b^{\frac{1}{2}}} \sup_{\lambda\in(0,\,1]} \frac{\left|1-\frac{(-1)^\nu}{2\nu+1}\frac{T_{2\nu+1}(\sqrt{\lambda})}{\sqrt{\lambda}}\right|}{\sqrt{\lambda}}.$$

Clearly, the constant

$$C_\nu = \sup_{\lambda\in(0,\,1]} \frac{\left|1-\frac{(-1)^\nu}{2\nu+1}\frac{T_{2\nu+1}(\sqrt{\lambda})}{\sqrt{\lambda}}\right|}{\sqrt{\lambda}}, \tag{3.5}$$

is independent of b.

□

The following two results are proven in the Appendix.

Proposition 3.4. *The constant C_ν, (3.5), defined in Proposition 3.3, satisfies the bound*

$$\frac{C_\nu}{2\nu+1} \le 2.$$

Proposition 3.5. *Define the polynomial*

$$s_\nu(t) = (-1)^\nu \frac{1}{2\nu+1}\frac{T_{2\nu+1}(\sqrt{t})}{\sqrt{t}}. \tag{3.6}$$

Then, the following estimate holds for $t \in [0,1]$

$$\left(1-(2\nu+1)^2 s_\nu^2(t)t\right) s_\nu(t) = \left(1-T_{2\nu+1}^2(\sqrt{t})\right) s_\nu(t) \ge -1+\delta_0,$$

with $\delta_0 = 1-\frac{2}{3\sqrt{3}} \in (0,1)$.

4. Main assumptions

The main assumption in our analysis is that a "*weak approximation property*" is satisfied by certain coarse spaces. For the scalar problems, (2.1), considered here, the relevant coarse spaces consist of piecewise constant vectors. Namely, it is well-known that a finite element function, v, can be approximated by a piecewise constant function $I_H v$ in L_2. The latter is defined based on sets $\widetilde{\mathcal{A}}_i$ (subdomains of Ω formed as the union of fine–grid elements that cover our composite aggregates, which we later do not distinguish) with diameter $\mathcal{O}(H)$. Over each set $\widetilde{\mathcal{A}}_i$, $I_H v$ is constant, for example equal to the average value of v over an extended domain Ω_i of $\widetilde{\mathcal{A}}_i$ with comparable diameter $\mathcal{O}(H)$ (for details see [SA]). Then, for A as in (2.2), i.e., obtained by a finite element discretization of the model problem (2.1), the following is a standard estimate in L_2 in terms of the energy norm $\|.\|_a \simeq |\,.\,|_{H^1(\Omega)}$, corresponding to continuous problem (2.1),(see [SA])

$$\|v - I_H v\|_0 \le c_a\, H\, \|v\|_a, \tag{4.1}$$

where

$$H = \max_i \operatorname{diam}(\widetilde{\mathcal{A}}_i). \tag{4.2}$$

Rewriting this in terms of vectors leads to the following estimate (with a different constant c_a)

$$h^{\frac{d}{2}} \, \|\mathbf{v} - \underline{I_H v}\| \leq c_a H \, \|\mathbf{v}\|_A,$$

where $d = 2$ or $d = 3$ is the dimension of the domain Ω where the (Laplacian–like) PDE of interest (2.1) is posed. Since the finite element discretization for such problems satisfies, $\|A\| \simeq h^{d-2}$, we arrive at

$$\|\mathbf{v} - \underline{I_H v}\| \leq c_a \frac{H}{h} \, \frac{1}{\|A\|^{\frac{1}{2}}} \, \|\mathbf{v}\|_A.$$

In the analysis of SA to follow, we will have $\frac{H}{h} \simeq (2\nu+1)^{k+1}$, where $\nu \geq 1$ is the degree of the polynomial used to define the prolongator in (2.3)-(2.4). Also, here $H = H_k$ defined in (4.2), is the characteristic diameter of the composite aggregates at coarsening level k.

We summarize this weak approximation estimate as our main assumption. Let $Q_k : \mathbb{R}^{n_0} \mapsto \mathbb{R}^{n_k}, k = 0, \ldots, \ell$ be the restriction mappings such that $I_k Q_k$ are ℓ_2-projections onto Range(I_k). The following approximation property is our main assumption:

$$\|\mathbf{v} - I_{k+1} Q_{k+1} \mathbf{v}\| \leq \sigma_a \, \frac{(2\nu+1)^{k+1}}{b_0^{\frac{1}{2}}} \, \|\mathbf{v}\|_A, \tag{4.3}$$

where $b = b_0$ is an available upper bound of $\|A\|$. The weak approximation property (4.3) is satisfied under the assumption that the diameter of composite aggregates, $\widetilde{\mathcal{A}}_i^k$, satisfies the estimate

$$\max_{i \in \mathcal{N}_k} \operatorname{diam}(\widetilde{\mathcal{A}}_i^k) \leq C_{\mathcal{A}} \, (2\nu+1)^{k+1} h. \tag{4.4}$$

A typical choice in practice is $\nu = 1$, but values $\nu > 1$ may be useful when a very aggressive coarsening is performed.

We conclude this section with several remarks.

Remark 4.1. *We comment that the constant σ_a (or the constant c_a in the Poincaré inequality* (4.1)*) generally depends on the shape of aggregates. A simple way to overcome that dependence, is to prove local estimates, composite aggregate–by-composite aggregate, by embedding each of them into a regularly shaped domain Ω_i (such as ball or box). Then, however, the overall constant σ_a, will depend on the maximum number of overlaps of the subdomains Ω_i. (For details, see* [SA]*.) For the specific general aggregation procedure that we describe in Section 7, it is easily seen that σ_a can be kept under control, uniformly, with respect to the number of coarsening levels used.*

Remark 4.2. *As with any algebraic multigrid method, to achieve a method with good overall storage and computational cost, it is important that the operators of the constructed multigrid hierarchy be sparse. For SA methods, this amounts to balancing the construction of the aggregates and the selection of the prolongation smoothing. In the last section, we comment on one way to define aggregates that guarantees uniform sparsity pattern of the coarse level matrices, which makes it easier to control the complexity of the overall method.*

Remark 4.3. *Our proof below allows for polynomial degree ν that can depend on the level index, i.e., $\nu = \nu_k$, that may be employed to optimize the complexity of the resulting SA method.*

5. Analysis details

We recall the definition (2.9) of the symmetrized smoother $\overline{M}_k$ coming from any A_k–convergent smoother M_k. We begin with the minimal assumption on smoother M_k (2.7) that implies that M_k provides an A_k–convergent iteration. As already mentioned, if M_k is spectrally equivalent to the Richardson one, $\|A_k\|\, I$ and properly scaled as

$$\mathbf{v}^T A_k \mathbf{v} \le \|A_k\| \mathbf{v}^T \mathbf{v} \le \mathbf{v}^T M_k \mathbf{v}, \tag{5.1}$$

then (2.7) is trivially satisfied. For example, we can let $M_k = b_k\, I$ where b_k is any available upper bound of $\|A_k\|$ (see (5.2)).

We note that for the SA method with aggressive coarsening, i.e., when we consider values $\nu > 1$, we later utilize a more specialized smoother that satisfies, in addition to (2.7), also an estimate of the form (2.8).

Preliminary estimates. We begin with a lemma that is a modification of a similar result in [SA].

Lemma 5.1. *Let $b_0 \ge \|A\|$ be an explicitly available upper bound of $\|A\|$. We set*

$$b_k = \frac{b_0}{(2\nu+1)^{2k}}, \quad k = 0, \dots, \ell. \tag{5.2}$$

Further, let $S_k = \varphi_\nu(A_k) = s_\nu\left(\frac{1}{b_k}\, A_k\right)$ (see (3.6)) with $b = b_k$ for all $k = 0, \dots, \ell - 1$. Then for every $k = 0, \dots, \ell$ it holds that

$$\|A_{k+1}\| \le \|A_k S_k^2\| \le b_{k+1}.$$

Proof. Assume $\|A_k\| \le b_k$ for some $k < \ell$. Recall that $(I_{k+1}^k)^T I_{k+1}^k = I$. Then, since $P_{k+1}^k = S_k I_{k+1}^k$, we have

$$\|A_{k+1}\| = \sup_{\mathbf{v}\in\mathbb{R}^{n_{k+1}}} \frac{\mathbf{v}^T A_{k+1}\mathbf{v}}{\mathbf{v}^T\mathbf{v}} = \sup_{\mathbf{v}\in\mathbb{R}^{n_{k+1}}} \frac{\mathbf{v}^T (I_{k+1}^k)_k^T S_k^T A_k S_k I_{k+1}^k \mathbf{v}}{(I_{k+1}^k\mathbf{v})^T (I_{k+1}^k\mathbf{v})} \le \sup_{\mathbf{v}\in\mathbb{R}^{n_k}} \frac{\mathbf{v}^T S_k^T A_k S_k \mathbf{v}}{\mathbf{v}^T\mathbf{v}}.$$

Further, using the fact that $S_k = \varphi_\nu(A_k)$ with $b = b_k$ and the property (3.2) of φ_ν, we get,

$$\begin{aligned} \mathbf{v}^T S_k^T A_k S_k \mathbf{v} &= \mathbf{v}^T \varphi_\nu^2(A) A \mathbf{v} \le \sup_{t\in\sigma(A_k)} \varphi_\nu(t)^2 t \|\mathbf{v}\|^2 \\ &\le \sup_{t\in[0,b_k]} \varphi_\nu(t)^2 t\, \|\mathbf{v}\|^2 \le \tfrac{b_k}{(2\nu+1)^2}\|\mathbf{v}\|^2. \end{aligned}$$

That is,

$$\|A_{k+1}\| \le \frac{b_k}{(2\nu+1)^2} = b_{k+1},$$

and the proof is complete by induction. □

Remark 5.1. *We point out that, we have the freedom to perturb the upper bounds b_k by a factor $1+\delta_k$ thus leading to the recurrence of the perturbed upper bounds $\widetilde{b}_k$,*

$$\widetilde{b}_{k+1} = \widetilde{b}_k(1+\delta_{k+1}) = b_k \prod_{j\le k+1}(1+\delta_j).$$

The analysis to follow remains valid with b_k replaced by $\widetilde{b}_k$ as long as the product $\prod_{j\le\ell}(1+\delta_j)$ stays uniformly bounded. The latter holds, for example, if $\delta_j \le C_0 j^{-2}$. The fact that we can perturb b_k allows us to assume that the matrix $s_\nu(b_k^{-1}A_k)$ is invertible. Note that $s_\nu(t)$ admits the factorization

$$s_\nu(t) = \prod_j (1-\tau_j^{-1}t), \quad \tau_j = \sin^2\Big(\frac{j}{2\nu+1}\,\pi\Big),\ j=1,\ \ldots,\ \nu.$$

Then, if the eigenvalues λ_s of A_k are such that $b_k^{-1}\lambda_s \neq \tau_j$ for all j, then each factor $I-\tau_j^{-1}b_k^{-1}A_k$ is an invertible matrix and hence so is $s_\nu(b_k^{-1}A_k)$. Otherwise with the perturbation $1+\delta_k$, we can ensure (in theory) that this does not happen. For example, if λ_s is such that $0<\tau_0 \equiv \min\{\frac{1}{2}b_k^{-1}\lambda_{\min},\ \tau_1\} \le \tau_{j_s-1} < b_k^{-1}\lambda_s \le \tau_{j_s}$, with the perturbation $1+\delta_k$, we can ensure $\tau_{j_s-1} < \frac{1}{1+\delta_k}b_k^{-1}\lambda_s < \tau_{j_s}$ if

$$\delta_k < \min_s \frac{b_k^{-1}\lambda_s - \tau_{j_s-1}}{\tau_{j_s-1}}.$$

Having ensured that $S_k = s_\nu(b_k^{-1}A_k)$ is invertible, implies that the smoothed prolongator $P_{k+1}^k = S_k I_{k+1}^k$ has the same column rank as I_{k+1}^k and since I_{k+1}^k has full column rank (it is actually orthogonal matrix), then so has P_{k+1}^k. In conclusion, starting with s.p.d. matrix $A_0 = A$ all coarse-level matrices A_k will stay positive definite. This is what we assume in the analysis to follow to avoid some technical details.

Next, we formulate the XZ-identity ([XZ]) in its matrix-vector form suitable for our analysis.

Given multigrid smoothers defined by M_k such that $M_k^T + M_k - A_k$ is coercive in terms of A_k, giving rise to the symmetrized smoother $\overline{M}_k$ (see (2.9)), interpolation matrices P_{k+1}^k, and the coarse matrices defined as $A_{k+1} = (P_{k+1}^k)^T A_k P_{k+1}^k$, the following main XZ-identity holds (cf., [Va08]):

$$\begin{aligned}\mathbf{v}^T A\mathbf{v} \le \mathbf{v}^T B\mathbf{v} \ &= \inf_{(\mathbf{v}_k)} \Big[\mathbf{v}_\ell^T A_\ell \mathbf{v}_\ell + \sum_{k<\ell}\Big(\mathbf{v}_k^f + M_k^{-T}A_k P_{k+1}^k \mathbf{v}_{k+1}\Big)^T \\ &\qquad \overline{M}_k\Big(\mathbf{v}_k^f + M_k^{-T}A_k P_{k+1}^k \mathbf{v}_{k+1}\Big)\Big].\end{aligned} \tag{5.3}$$

The inf here is taken over the components $(\mathbf{v}_k)$ of all possible decompositions of $\mathbf{v}$ obtained as follows:

(i) Starting with $\mathbf{v}_0 = \mathbf{v}$,
(ii) for $k \ge 0$, we let $\mathbf{v}_k = \mathbf{v}_k^f + P_{k+1}^k \mathbf{v}_{k+1}$, i.e., first choosing $\mathbf{v}_{k+1} \in \mathbb{R}^{n_{k+1}}$ arbitrary, we then let $\mathbf{v}_k^f = \mathbf{v}_k - P_{k+1}^k \mathbf{v}_{k+1}$.

We recall here that $Q_k : \mathbb{R}^{n_0} \to \mathbb{R}^{n_k}$ are linear mappings such that $I_k Q_k$ are ℓ_2 projections onto Range(I_k), that is,

$$Q_k = ((I_k)^T I_k)^{-1}(I_k)^T = (I_k)^T.$$

We consider the following particular decomposition

$$\mathbf{v}_k^f = (Q_k - P_{k+1}^k Q_{k+1})\mathbf{v}, \quad \mathbf{v}_{k+1} = Q_{k+1}\mathbf{v},$$

and observe that applying Cauchy–Schwarz inequality (in the $\overline{M}_k$–inner product) to the sum in (5.3) results in the estimate

$$\begin{aligned}
&\sum_{k<l}(\mathbf{v}_k^f + M_k^{-T}A_k P_{k+1}^k \mathbf{v}_{k+1})^T \overline{M}_k (\mathbf{v}_k^f + M_k^{-T}A_k P_{k+1}^k \mathbf{v}_{k+1}) \\
&\le 2\sum_{k<l}(\mathbf{v}_k^f)^T \overline{M}_k \mathbf{v}_k^f + 2\sum_{k<l}(P_{k+1}^k \mathbf{v}_{k+1})^T A_k M_k^{-1}\overline{M}_k M_k^{-T} A_k P_{k+1}^k \mathbf{v}_{k+1}) \\
&\le 2\sum_{k<l}(\mathbf{v}_k^f)^T \overline{M}_k \mathbf{v}_k^f + \tfrac{2}{\alpha}\sum_{k<l}(P_{k+1}^k \mathbf{v}_{k+1})^T A_k P_{k+1}^k \mathbf{v}_{k+1} \\
&= 2\sum_{k<l}(\mathbf{v}_k^f)^T \overline{M}_k \mathbf{v}_k^f + \tfrac{2}{\alpha}\sum_{k<l}(P_{k+1}^k \mathbf{v}_{k+1})^T A_k P_{k+1}^k \mathbf{v}_{k+1} \\
&= 2\sum_{k<l}(\mathbf{v}_k^f)^T \overline{M}_k \mathbf{v}_k^f + \tfrac{2}{\alpha}\sum_{k<l}(\mathbf{v}_{k+1})^T A_{k+1} \mathbf{v}_{k+1}.
\end{aligned} \tag{5.4}$$

From here, we see that in order to bound the relative condition number of the V–cycle preconditioner B with respect to A, (due to estimate (5.3)), based on our choice of the smoother as in (2.7), it is sufficient to bound the expressions (i) and (ii) below:

(i) $$\sum_{k<\ell}(\mathbf{v}_k^f)^T \overline{M}_k \mathbf{v}_k^f = \sum_{k<\ell}\left((Q_k - P_{k+1}^k Q_{k+1})\mathbf{v}\right)^T \overline{M}_k \left((Q_k - P_{k+1}^k Q_{k+1})\mathbf{v}\right)$$

and

(ii) $$\sum_{k\le\ell}\mathbf{v}_k^T A_k \mathbf{v}_k = \sum_{k\le\ell}\mathbf{v}^T Q_k^T A_k Q_k \mathbf{v},$$

both in terms of $\mathbf{v}^T A\mathbf{v}$.

Estimating the first sum (i). Recall that $P_{k+1}^k = S_k I_{k+1}^k$, $S_k = \varphi_\nu(A_k)$ with $b = b_k \ge \|A_k\|$ given by (5.2), $I_k = I_1^0 I_2^1 \dots I_k^{k-1}$ and $(I_k^{k-1})^T I_k^{k-1} = I$. Note that (see (3.3)) $\|S_k\| \le \sup_{t\in[0,\, b_k]} |\varphi_\nu(t)| = 1$. We start with the inequality,

$$\begin{aligned}
\|\left(Q_k - P_{k+1}^k Q_{k+1}\right)\mathbf{v}\| &= \|\left(Q_k - S_k I_{k+1}^k Q_{k+1}\right)\mathbf{v}\| \\
&= \|S_k\left(Q_k - I_{k+1}^k Q_{k+1}\right)\mathbf{v} + (I - S_k)Q_k\mathbf{v}\| \\
&\le \|S_k\left(Q_k - I_{k+1}^k Q_{k+1}\right)\mathbf{v}\| + \|(I - S_k)Q_k\mathbf{v}\| \\
&\le \|S_k\|\|\left(Q_k - I_{k+1}^k Q_{k+1}\right)\mathbf{v}\| + \|(I - S_k)Q_k\mathbf{v}\| \\
&\le \|I_k\left(Q_k - I_{k+1}^k Q_{k+1}\right)\mathbf{v}\| + \|(I - S_k)Q_k\mathbf{v}\|.
\end{aligned}$$

Notice that

$$I - S_k = I - \varphi_\nu(A_k) = A_k^{-\frac{1}{2}}\left(I - \varphi_\nu(A_k)\right)A_k^{\frac{1}{2}}.$$

Based on estimate (3.4), we then get

$$\begin{aligned}
\|(I - S_k)Q_k\mathbf{v}\| &= \|A_k^{-\frac{1}{2}}\left(I - \varphi_\nu(A_k)\right)A_k^{\frac{1}{2}}Q_k\mathbf{v}\| \\
&\le \|A_k^{-\frac{1}{2}}\left(I - \varphi_\nu(A_k)\right)\|\|A_k^{\frac{1}{2}}Q_k\mathbf{v}\| \\
&\le \max_{t\in\sigma(A_k)} \tfrac{1-\varphi_\nu(t)}{\sqrt{t}}\ \|Q_k\mathbf{v}\|_{A_k} \\
&\le \sup_{t\in(0,\, b_k]} \tfrac{1-\varphi_\nu(t)}{\sqrt{t}}\ \|Q_k\mathbf{v}\|_{A_k} \\
&\le C_\nu \tfrac{1}{\sqrt{b_k}}\ \|Q_k\mathbf{v}\|_{A_k}
\end{aligned} \tag{5.5}$$

Thus, we arrived at the estimate

$$\|(Q_k - P_{k+1}^k Q_{k+1})\mathbf{v}\| \le \| \left(I_k Q_k - I_{k+1}Q_{k+1}\right)\mathbf{v}\| + \frac{C_\nu}{b_k^{1/2}} \|Q_k\mathbf{v}\|_{A_k}. \tag{5.6}$$

The final bound of sum (i) will be derived after we estimate sum (ii).

Estimating the second sum (ii). We bound next $\|Q_k\mathbf{v}\|_{A_k}$. Since $S_k = \varphi_\nu(A_k)$ with $b = b_k$, (3.3) gives $\|S_k\|_{A_k} \le 1$. Then, based on Lemma 5.1 and the property (3.2), we obtain

$$\begin{aligned}
\|Q_{k+1}\mathbf{v}\|_{A_{k+1}} &= \|P_{k+1}^k Q_{k+1}\mathbf{v}\|_{A_k} = \|S_k I_{k+1}^k Q_{k+1}\mathbf{v}\|_{A_k} \\
&= \|S_k(I_{k+1}^k Q_{k+1} - Q_k)\mathbf{v} + S_k Q_k\mathbf{v}\|_{A_k} \\
&\le \|S_k\left(I_{k+1}^k Q_{k+1} - Q_k\right)\mathbf{v}\|_{A_k} + \|S_k Q_k\mathbf{v}\|_{A_k} \\
&\le \|S_k\left(I_{k+1}^k Q_{k+1} - Q_k\right)\mathbf{v}\|_{A_k} + \|S_k\|_{A_k}\|Q_k\mathbf{v}\|_{A_k} \\
&\le \|A_k^{\frac{1}{2}} S_k(Q_k - I_{k+1}^k Q_{k+1})\mathbf{v}\| + \|Q_k\mathbf{v}\|_{A_k} \\
&\le \|A_k^{\frac{1}{2}} S_k\|\|(Q_k - I_{k+1}^k Q_{k+1})\mathbf{v}\| + \|Q_k\mathbf{v}\|_{A_k} \\
&= \|A_k^{\frac{1}{2}} \varphi_\nu(A_k)\|\|(Q_k - I_{k+1}^k Q_{k+1})\mathbf{v}\| + \|Q_k\mathbf{v}\|_{A_k} \\
&\le \tfrac{b_k^{1/2}}{(2\nu+1)}\|(Q_k - I_{k+1}^k Q_{k+1})\mathbf{v}\| + \|Q_k\mathbf{v}\|_{A_k} \\
&= \tfrac{b_0^{\frac{1}{2}}}{(2\nu+1)^{k+1}}\|I_k(Q_k - I_{k+1}^k Q_{k+1})\mathbf{v}\| + \|Q_k\mathbf{v}\|_{A_k} \\
&= \tfrac{b_0^{\frac{1}{2}}}{(2\nu+1)^{k+1}}\|(I_k Q_k - I_{k+1}Q_{k+1})\mathbf{v}\| + \|Q_k\mathbf{v}\|_{A_k} \\
&= b_{k+1}^{\frac{1}{2}}\|(I_k Q_k - I_{k+1}Q_{k+1})\mathbf{v}\| + \|Q_k\mathbf{v}\|_{A_k}.
\end{aligned} \tag{5.7}$$

Since $I_k Q_k$ is an ℓ_2-orthogonal projection onto Range(I_k) and $I_{k+1}Q_{k+1}$ is an ℓ_2-orthogonal projection onto Range$(I_{k+1}) \subset$ Range(I_k), we have

$$(I_k Q_k)^2 = I_k Q_k \text{ and } I_k Q_k I_{k+1} Q_{k+1} = I_{k+1} Q_{k+1},$$

therefore,

$$\left((I - I_k Q_k)\mathbf{u}\right)^T \left(I_k Q_k - I_{k+1}Q_{k+1}\right)\mathbf{v} = \mathbf{u}^T(I - I_k Q_k)(I_k Q_k - I_{k+1}Q_{k+1})\mathbf{v} = 0 \quad \forall \mathbf{u}, \mathbf{v},$$

that is, spaces Range$(I - I_k Q_k)$ and Range$(I_k Q_k - I_{k+1}Q_{k+1})$ are ℓ_2-orthogonal, and we have the Pythagorean theorem

$$\|\mathbf{v} - I_{k+1}Q_{k+1}\mathbf{v}\|^2 = \| \left(I_k Q_k - I_{k+1}Q_{k+1}\right)\mathbf{v}\|^2 + \|\mathbf{v} - I_k Q_k\mathbf{v}\|^2.$$

Therefore,

$$\| \left(I_k Q_k - I_{k+1}Q_{k+1}\right)\mathbf{v}\| \le \|\mathbf{v} - I_{k+1}Q_{k+1}\mathbf{v}\|.$$

That is, if we bound $\|\mathbf{v} - I_{k+1}Q_{k+1}\mathbf{v}\|$, the result will follow.

Use now the main estimate (4.3) which was our main assumption. It reads,

$$\|\mathbf{v} - I_k Q_k\mathbf{v}\|^2 \le \sigma_a^2 \frac{(2\nu+1)^{2k}}{b_0}\mathbf{v}^T A\mathbf{v},$$

where b_0 is an available upper bound of $\|A\|$. Then,

$$\| \left(I_k Q_k - I_{k+1}Q_{k+1}\right)\mathbf{v}\| \le \|\mathbf{v} - I_{k+1}Q_{k+1}\mathbf{v}\| \le \sigma_a \frac{(2\nu+1)^{k+1}}{b_0^{1/2}}\|\mathbf{v}\|_A = \sigma_a b_{k+1}^{-1/2}\|\mathbf{v}\|_A. \tag{5.8}$$

Substituting the latter estimate in (5.7) leads to the following recursive estimate,

$$\|Q_{k+1}\mathbf{v}\|_{A_{k+1}} \leq \|Q_k\mathbf{v}\|_{A_k} + \sigma_a\|\mathbf{v}\|_A.$$

That is, we proved the following main estimate

$$\|Q_k\mathbf{v}\|_{A_k} \leq (1+\sigma_a k)\ \|\mathbf{v}\|_A \leq (1+\sigma_a\ell)\|\mathbf{v}\|_A. \tag{5.9}$$

Thus the second sum is bounded as follows

$$\sum_{l\leq\ell} \mathbf{v}_k^T A_k \mathbf{v}_k = \sum_{k\leq\ell} \|Q_k\mathbf{v}\|_{A_k}^2 \leq (1+\sigma_a\ell)^2\ell\ \mathbf{v}^T A\mathbf{v}. \tag{5.10}$$

Note that the above estimate is independent of ν, that is, the stability of Q_k in energy (5.10) holds independently of the coarsening ratio (between two consecutive levels) $H_{k+1}/H_k = 2\nu+1$ even when ν gets large.

Completing the bound of the first sum (i). The estimate (5.6), together with (5.8) and (5.9), imply

$$\begin{aligned} \|\left(Q_k - P_{k+1}^k Q_{k+1}\right)\mathbf{v}\| &\leq \|\left(I_k Q_k - I_{k+1}Q_{k+1}\right)\mathbf{v}\| + \tfrac{C_\nu}{b_k^{1/2}}\ \|Q_k\mathbf{v}\|_{A_k} \\ &\leq \sigma_a \tfrac{(2\nu+1)^{k+1}}{b_0^{1/2}}\|\mathbf{v}\|_A + \tfrac{C_\nu}{b_k^{1/2}}\ (1+\sigma_a k)\ \|\mathbf{v}\|_A \\ &\leq \tfrac{\sigma_a(2\nu+1)+C_\nu(1+\sigma_a k)}{b_k^{1/2}}\|\mathbf{v}\|_A. \end{aligned} \tag{5.11}$$

Thus, we proved a weak approximation property for the *smoothed* prolongators. Note that the above bound depends on ν and grows linearly with increasing value of ν. This reflects the simple fact that the coarse-space, as it becomes smaller, loses its approximation capability.

Next, we prove two bounds for sum (i) depending on the choice of smoother M_k.

First consider the simple choice of $M_k \simeq b_k I$, i.e., spectrally equivalent to Richardson. Recall that $b_k \geq \|A_k\|$ is given by (5.2). To establish the bound of sum (i) in this case, we use the estimate

$$\begin{aligned} \|\overline{M}_k^{1/2}\left(Q_k - P_{k+1}^k Q_{k+1}\right)\mathbf{v}\| &\simeq b_k^{1/2}\|\left(Q_k - P_{k+1}^k Q_{k+1}\right)\mathbf{v}\| \\ &\leq \left(\sigma_a(2\nu+1) + C_\nu(1+\sigma_a k)\right)\|\mathbf{v}\|_A \\ &\leq (2\nu+1)\left(\sigma_a + 2(1+\sigma_a k)\right)\|\mathbf{v}\|_A \end{aligned} \tag{5.12}$$

In the last inequality, we used Proposition 3.4. This bound grows with $\nu \mapsto \infty$. Thus, simple smoothers spectrally equivalent to Richardson are appropriate only for bounded coarsening ratio (or bounded ν).

For the case of large ν, we consider an alternative smoother, M_k, that in addition to (2.7), satisfies also estimate (2.8), namely,

$$\mathbf{v}_k^T\overline{M}_k\mathbf{v}_k \leq \beta\left(b_{k+1}\ \|\mathbf{v}_k\|^2 + \mathbf{v}_k^T A_k\mathbf{v}_k\right). \tag{5.13}$$

Such smoother will be introduced and analyzed in the following section.

Using (5.13) for $\mathbf{v}_k^f = \left(Q_k - P_{k+1}^k Q_{k+1}\right)\mathbf{v}$, we obtain

$$\|\left(Q_k - P_{k+1}^k Q_{k+1}\right)\mathbf{v}\|_{\overline{M}_k}^2 \leq \beta\ \left(b_{k+1}\ \|\left(Q_k - P_{k+1}^k Q_{k+1}\right)\mathbf{v}\|^2 + \|\left(Q_k - P_{k+1}^k Q_{k+1}\right)\mathbf{v}\|_{A_k}^2\right).$$

Using the first line of estimate (5.11), together with (5.8) and (5.9), we have

(5.14)
$$\begin{aligned} b_{k+1}^{\frac{1}{2}} \| \left(Q_k - P_{k+1}^k Q_{k+1}\right) \mathbf{v}\| &\leq b_{k+1}^{\frac{1}{2}} \|(I_k Q_k - I_{k+1} Q_{k+1})\mathbf{v}\| + C_\nu \sqrt{\tfrac{b_{k+1}}{b_k}}\, \|Q_k \mathbf{v}\|_{A_k} \\ &\leq \left(\sigma_a + (1+\sigma_a \ell)\, \sup_\nu \tfrac{C_\nu}{2\nu+1}\right)\, \|\mathbf{v}\|_A. \end{aligned}$$

To bound the energy norm of $(Q_k - P_{k+1}^k Q_{k+1})\mathbf{v} = (Q_k - S_k I_{k+1}^k Q_{k+1})\mathbf{v}$ we use (5.9). Then,

$$\begin{aligned} \|(Q_k - S_k I_{k+1}^k Q_{k+1})\mathbf{v}\|_{A_k} &\leq \|Q_k \mathbf{v}\|_{A_k} + \|S_k I_{k+1}^k Q_{k+1}\mathbf{v}\|_{A_k} \\ &\leq \|Q_k \mathbf{v}\|_{A_k} + \|Q_{k+1}\mathbf{v}\|_{A_{k+1}} \\ &\leq 2(1+\sigma_a \ell)\|\mathbf{v}\|_A. \end{aligned} \tag{5.15}$$

Combining (5.14) and (5.15), (5.13) and using Proposition 3.4 yields the final desired bound

(5.16)
$$\begin{aligned} \| \left(Q_k - P_{k+1}^k Q_{k+1}\right) \mathbf{v}\|_{\overline{M}_k}^2 &\leq \beta \left(\left(\sigma_a + (1+\sigma_a \ell)\, \sup_\nu \tfrac{C_\nu}{2\nu+1}\right)^2 + 4(1+\sigma_a \ell)^2 \right) \|\mathbf{v}\|_A^2 \\ &\leq \beta \left[\left(\sigma_a + 2(1+\sigma_a \ell)\right)^2 + 4(1+\sigma_a \ell)^2 \right] \|\mathbf{v}\|_A^2. \end{aligned}$$

Final estimates. In conclusion, we are ready to complete the proof of the following main result (given for $\nu = 1$ and Richardson smoother in [SA]).

Theorem 5.1. *Consider the s.p.d. matrix A as in (2.2) that is a finite element discretization of the model PDE (2.1). We define the SA AMG as described in Section 2 and Section 3.*

We make the following assumptions:

- *The approximation property (4.3) for the composite piecewise constant interpolants I_k (from coarse level $k+1$ all the way up to finest level 0) holds. This is the case under assumption (4.4) that the kth level composite aggregates have diameter that grows not faster than $(2\nu+1)^{k+1}h$ (where h is the finest level meshsize).*
- *The smoothed prolongation operators are defined as $P_{k+1}^k = \varphi_\nu(A_k) I_{k+1}^k$, where I_{k+1}^k is the scaled piecewise constant interpolant from coarse level $k+1$ to the next fine level k, and the prolongation smoothers, φ_ν, are defined by (3.1), with $b = b_k \geq \|A_k\|$, and b_k given by (5.2) with b_0 being an explicitly available upper bound of $\|A\|$ (the original s.p.d. matrix).*
- *The multigrid smoother has error propagation operator, $I - M_k^{-1} A_k$, where either $M_k \simeq b_k I$, with b_k given by (5.2) or a more general smoother that satisfies (2.7).*
- *In the case of aggressive coarsening, i.e., for large values of ν, in addition to property (2.7) we assume that the symmetrized smoother (defined in (2.9)) satisfies estimate (2.8).*

Then the resulting $V(1,1)$–cycle MG preconditioner, B, is nearly spectrally equivalent to A with $K_* \leq C(\ell^3 + C_0 \nu^2 \ell^3)$, where K_* is the constant in (2.10). The constant C_0 is positive for smoothers spectrally equivalent to Richardson. That is K_* grows quadratically with the coarsening factor (or equivalently quadratically with ν). In the case of smoother with property (2.8) $C_0 = 0$, i.e., the constant K_* is bounded independently of $\nu \mapsto \infty$.

Proof. The proof, in both cases (Richardson smoother or smoother with property (2.8)) exploits the already derived bounds of the sums (i) and (ii). In the case of smoother spectrally equivalent to Richardson, we use estimates (5.12) and (5.10), whereas in the second case (of smoother with property (2.8)) we use estimates (5.16) and (5.15).

For an illustration, we provide the details for the Richardson smoother. We use the particular decomposition $\mathbf{v}_k^f = (Q_k - P_{k+1}^k Q_{k+1})\mathbf{v}$ and $\mathbf{v}_{k+1} = Q_{k+1}\mathbf{v}$ in the XZ-identity. We first notice, that for the Richardson smoother $M_k = b_k I$, we have

$$(\mathbf{v}_k^f)^T \overline{M}_k \mathbf{v}_k^f \leq b_k^2 \, (\mathbf{v}_k^f)^T \, (2b_k I - A_k)^{-1} \, \mathbf{v}_k^f \leq b_k \, \|\mathbf{v}_k^f\|^2.$$

Similarly, we have

$$\begin{aligned} \mathbf{v}_k^T A_k (M_k + M_k^T - A_k)^{-1} A_k \mathbf{v}_k \; &= \mathbf{v}_k^T A_k (2b_k \, I - A_k)^{-1} A_k \mathbf{v}_k \\ &\leq \tfrac{1}{b_k} \, \|A_k \mathbf{v}_k\|^2 \leq \mathbf{v}_k^T A_k \mathbf{v}_k, \end{aligned}$$

which verifies property (2.7) (with $\alpha = 1$).

Using this inequality in the upper bound (5.4), for the XZ-identity (5.3) and estimates (5.12) and (5.10), we obtain

$$\begin{aligned} \mathbf{v}^T B \mathbf{v} \; &\leq \left[\|Q_\ell \mathbf{v}\|_{A_\ell}^2 + 2 \sum_{k<l} b_k \| \left(Q_k - P_{k+1}^k Q_{k+1}\right) \mathbf{v}\|^2 + 2 \sum_{k<l} \|Q_{k+1}\mathbf{v}\|_{A_{k+1}}^2 \right] \\ &\leq \|Q_\ell \mathbf{v}\|_{A_\ell}^2 + 2 \sum_{k \leq \ell} \|Q_k \mathbf{v}\|_{A_k}^2 + 2 \sum_{k<\ell} b_k \| \left(Q_k - P_{k+1}^k Q_{k+1}\right) \mathbf{v}\|^2 \\ &\leq C \left[\ell^3 + (2\nu+1)^2 \sum_{k<\ell} k^2 \right] \|\mathbf{v}\|_A^2 \\ &\leq C \, (\ell^3 + (2\nu+1)^2 \ell^3) \, \|\mathbf{v}\|_A^2, \end{aligned}$$

where the constant, C, is independent of ν.

□

6. A polynomial smoother in the case of aggressive coarsening

In this section we introduce a special polynomial smoother M_k that satisfies the two properties we used in the analysis; namely, the coercivity estimate (2.7) and assumption (2.8).

The construction makes use of two properties of the polynomial φ_ν, (or its normalized form s_ν, (3.6)), as ν grows large, formulated earlier as Proposition 3.4 and Proposition 3.5. Their proofs are found in the Appendix.

We define the error propagation operator of both pre- and post- smoother of our symmetric multigrid cycle as

$$I - M_k^{-T} A_k = I - M_k^{-1} A_k = \left(I - \frac{1}{b_{k+1}} A_k S_k^2 \right) S_k. \tag{6.1}$$

We note that this expression does not necessarily define an invertible matrix M_k, only the product $X_k = M_k^{-T} A_k$ is well–defined and it gives rise to a positive semi–definite matrix. Nevertheless the iteration method corresponding to (6.1) can be implemented as a sequence of Jacobi sweeps. For this, we need to know the roots of the polynomials involved to factor them as a product of first-order polynomials $I - M_k^{-1} A_k = \prod_j (I - \tau_j^{-1} b_k^{-1} A_k)$. To implement the smoother M_k applied to

$$A_k \mathbf{u} = \mathbf{f},$$

we can use the following standard composite algorithm. Given initial approximation $\mathbf{u}_0$, for $j \geq 1$, we compute $\mathbf{u}_j$ from

$$\tau_j b_k(\mathbf{u}_j - \mathbf{u}_{j-1}) = \mathbf{f} - A_k \mathbf{u}_{j-1}.$$

The consecutive errors $\mathbf{e}_j = \mathbf{u} - \mathbf{u}_j$ satisfy the recurrence $\mathbf{e}_j = (I - \frac{1}{\tau_j} b_k^{-1} A_k)\mathbf{e}_{j-1}$, which shows that the above composite iteration does lead to the iteration matrix $I - M_k^{-1} A_k$.

Letting $S_k = \varphi_\nu(A_k) = s_\nu(b_k^{-1} A_k)$, i.e., using polynomial that is independent of the scaling b_k, recalling our choice of polynomial φ_ν (defined in (3.1) for $b = 1$, or see Proposition 3.5), we have

$$s_\nu(t) = (-1)^\nu \frac{1}{2\nu+1} \frac{T_{2\nu+1}(\sqrt{t})}{\sqrt{t}}.$$

This leads to the following equivalent definition of M_k,

$$\begin{aligned} I - M_k^{-1} A_k &= \left(I - \tfrac{1}{b_{k+1}} A_k S_k^2\right) S_k = \left(I - \tfrac{(2\nu+1)^2}{b_k} A_k S_k^2\right) S_k \\ &= \left(I - \left(T_{2\nu+1}\left(\sqrt{b_k^{-1} A_k}\right)\right)^2\right) s_\nu\left(b_k^{-1} A_k\right). \end{aligned} \tag{6.2}$$

The following factored representation can be derived readily based on properties of Chebyshev polynomials,

$$(I - T_{2\nu+1}^2(\sqrt{t}))s_\nu(t) = \prod_{j:\, 0 \leq j \leq 2\nu} \left(1 - \frac{1}{\cos^2\left(\frac{j}{2\nu+1}\Pi\right)} t\right) \prod_{j:\, 1 \leq j \leq \nu} \left(1 - \frac{1}{\sin^2\left(\frac{j}{2\nu+1}\,\Pi\right)} t\right).$$

Then, we easily find the roots τ_j.

We recall the definition of symmetrized smoother and its relation to the pre- and post-smoother:

$$(I - M_k^{-T} A_k)(I - M_k^{-1} A_k) = I - \overline{M}_k^{-1} A_k, \quad \overline{M}_k = M_k(M_k + M_k^T - A_k)^{-1} M_k^T.$$

Since

$$I - \overline{M}_k^{-1} A_k = (I - M_k^{-T} A_k)(I - M_k^{-1} A_k) = \left(I - \frac{1}{b_{k+1}} A_k S_k^2\right)^2 S_k^2,$$

we have

$$\overline{M}_k^{-1} = A_k^{-1}\left[I - \left(I - \frac{1}{b_{k+1}} A_k S_k^2\right)^2 S_k^2\right]. \tag{6.3}$$

It is clear that we can choose the number of smoothing steps to be proportional to the degree of the prolongator smoother used and, in turn, to the coarsening ratio. In practice, it may be convenient to choose this smoother to be tied to the prolongation smoother. In general, however, it is possible to use a value of ν in the definition of this smoother to differ from the value of ν we have used to define $P_{k+1}^k = s_\nu(b_k^{-1} A_k) I_{k+1}^k$, as long as these two numbers are comparable.

The following lemma verifies assumption (2.8) which was a key tool in proving convergence rate independent of the coarsening ratio (or parameter ν). It also shows that the direct definition of the symmetrized smoother (6.3) actually defines a s.p.d. matrix $\overline{M}_k^{-1}$ in the subspace $\mathrm{Null}^\perp(A_k)$ spanned by the eigenvectors of A_k with positive corresponding

eigenvalues. Based on Remark 5.1, we have that for the SA method we can assume that A_k is s.p.d., then the polynomial smoothers $\overline{M}_k$ is s.p.d. and as consequence M_k is seen to be invertible. For completeness, in what follows in the present section, we consider the general case assuming that A_k may potentially have a non–empty nullspace Null (A_k), that is, A_k being positive semi–definite (and symmetric).

Lemma 6.1. *Let $\{\mathbf{v}_i\}$ be eigenvectors of A_k and $\lambda_i(S_k)$ corresponding eigenvalues of S_k. We define*

$$\begin{aligned} U_1 &= \{span\{\mathbf{v}_i\} \,:\, |\lambda_i(S_k)| \le q\}\,, \\ U_2 &= \{span\{\mathbf{v}_i\} \,:\, |\lambda_i(S_k)| > q\}\,, \end{aligned}$$

where $q \in (0,1)$ is given parameter. Then, the right-hand side of (6.3) defines an invertible matrix in the subspace Null$^{\perp}$ $(A_k) = span\,\{\mathbf{v}_j :\ \lambda_j > 0\}$, which shows that $\overline{M}_k$ is s.p.d. restricted to that subspace. Also, $\overline{M}_k$ can be extended on $\mathbb{R}^{n_k}$ as semi–definite matrix vanishing on Null (A_k). Finally, the following estimates are valid:

$$\mathbf{x}^T\overline{M}_k\mathbf{x} \;\le\; \frac{1}{1-q^2}\,\mathbf{x}^T A_k\mathbf{x} \;for\; \mathbf{x}\in U_1, \tag{6.4}$$

$$\mathbf{x}^T\overline{M}_k\mathbf{x} \;\le\; \frac{b_{k+1}}{q^2}\|\mathbf{x}\|^2 \;for\; \mathbf{x}\in U_2. \tag{6.5}$$

In addition, for every $\mathbf{x}\in\mathbb{R}^{n_k}$ it holds that

$$\|\mathbf{x}\|^2_{\overline{M}_k} \le \frac{1}{1-q^2}\,\|\mathbf{x}\|^2_{A_k} + \frac{b_{k+1}}{q^2}\|\mathbf{x}\|^2. \tag{6.6}$$

Proof. Note that A_k, S_k and $I - b_{k+1}^{-1}A_kS_k^2$ are all symmetric, mutually commute, have common eigenvectors and U_1 and U_2 are their invariant subspaces. Lemma 5.1 gives $\varrho(S_k^2A_k) \le b_{k+1}$ and therefore $I - b_{k+1}^{-1}A_kS_k^2$ is positive semi-definite and $\varrho(I - b_{k+1}^{-1}A_kS_k^2) \le 1$. Further, (3.3) gives $\varrho(S_k) \le 1$.

For any given $\mathbf{x}\in$ Null$^{\perp}$ (A_k), use its expansion in terms of the orthonormal basis of eigenvectors $\{\mathbf{v}_j\}$ of the symmetric matrix A_k

$$\mathbf{x} = \sum_j c_j\mathbf{v}_j.$$

Then, denoting with s_j the eigenvalues of S_k (equal to $s_\nu(b_k^{-1}\lambda_j)$), using the fact that $\overline{M}_k^{-1}$ is well–defined as the r.h.s. of (6.3) when restricted to Null$^{\perp}$ (A_k), we have

$$\mathbf{x}^T\overline{M}_k^{-1}\mathbf{x} = \sum_j \lambda_j^{-1}\left(1 - \left(1 - \frac{1}{b_{k+1}}\lambda_j s_j^2\right)^2 s_j^2\right) c_j^2.$$

Using formula (6.2), we then have

$$b_k\,\mathbf{v}^T\overline{M}_k^{-1}\mathbf{x} = \sum_j \frac{b_k}{\lambda_j}\left(1 - \left(1 - T_{2\nu+1}^2\left(\sqrt{\frac{\lambda_j}{b_k}}\right)\right)^2 s_j^2\right) c_j^2.$$

We show next that the above expression is non-zero for any nonzero vector $\mathbf{x}$. Denote $t_j = \frac{\lambda_j}{b_k} \in [0,\ 1]$.

We partition the above sum into two parts. The first sum corresponds to values $t_j \in (0,\ \frac{1}{(2\nu+1)^2}]$. For such t_j, we have

$$\begin{aligned}
&t_j^{-1}\left(1-\left(1-T_{2\nu+1}^2(\sqrt{t_j})\right)^2 \tfrac{1}{t_j(2\nu+1)^2}T_{2\nu+1}^2(\sqrt{t_j})\right)\\
&= \tfrac{1}{t_j^2(2\nu+1)^2}\left(t_j(2\nu+1)^2-\left(1-T_{2\nu+1}^2(\sqrt{t_j})\right)^2 T_{2\nu+1}^2(\sqrt{t_j})\right)\\
&\geq \tfrac{1}{t_j}\left(t_j(2\nu+1)^2-\left(1-T_{2\nu+1}^2(\sqrt{t_j})\right)^2 T_{2\nu+1}^2(\sqrt{t_j})\right)\\
&\geq (2\nu+1)^2 - t_j^{-1}T_{2\nu+1}^2(\sqrt{t_j}) > 0.
\end{aligned}$$

In the last line we used the inequality $|T_{2\nu+1}(x)| < (2\nu+1)x$ for $x=\sqrt{t_j}$. The inequality is strict for $\nu \geq 1$ and $1 \geq x > 0$. Indeed, $T_3(x) = 4x^3 - 3x$ satisfies $-3x < T_3(x) < 3x$ for $x \in (0,1]$, that is, $|T_3(x)| < 3x$. Assuming by induction that $|T_{2\nu-1}(x)| < (2\nu-1)x$, we then have

$$|T_{2\nu+1}(x)| = |2xT_{2\nu}(x) - T_{2\nu-1}(x)| \leq 2x + |T_{2\nu-1}(x)| < 2x + (2\nu-1)x = (2\nu+1)x,$$

which proves the result.

Now consider the second case, i.e., $1 \geq t_j > \frac{1}{(2\nu+1)^2}$. We have

$$\begin{aligned}
&t_j^{-1}\left(1-\left(1-T_{2\nu+1}^2(\sqrt{t_j})\right)^2 \tfrac{1}{t_j(2\nu+1)^2}T_{2\nu+1}^2(\sqrt{t_j})\right)\\
&\geq 1-\left(1-T_{2\nu+1}^2(\sqrt{t_j})\right)^2 \frac{1}{t_j(2\nu+1)^2}T_{2\nu+1}^2(\sqrt{t_j})\\
&\geq 1-\left(1-T_{2\nu+1}^2(\sqrt{t_j})\right)^2 T_{2\nu+1}^2(\sqrt{t_j})\\
&\geq 1-\max_{x\in[0,1]} x^2(1-x) = \frac{23}{27}.
\end{aligned}$$

This completes the proof that $\overline{M}_k^{-1}$ is a well-defined s.p.d. matrix on $\text{Null}^{\perp}(A_k)$, for any $\nu \geq 1$.

Next, we prove the second part of the lemma, namely estimates (6.4),(6.5) and (6.6).

Let $\mathbf{x} \in U_1$ and $\mathbf{x} \in \text{Null}^{\perp}(A_k)$. It is clear that $A_k^{\frac{1}{2}}\mathbf{x}$ also belongs to U_1. We have

$$\begin{aligned}
\mathbf{x}^T\overline{M}_k\mathbf{x} &= \mathbf{x}^T A_k\left[I-\left(I-\frac{1}{b_{k+1}}A_kS_k^2\right)^2 S_k^2\right]^{-1}\mathbf{x}\\
&= \left(A_k^{\frac{1}{2}}\mathbf{x}\right)^T\left[I-\left(I-\frac{1}{b_{k+1}}A_kS_k^2\right)^2 S_k^2\right]^{-1}A_k^{\frac{1}{2}}\mathbf{x}\\
&= \left(A_k^{\frac{1}{2}}\mathbf{x}\right)^T\left(1-\left(1-T_{2\nu+1}^2(\left(\frac{1}{\sqrt{b_k}}A_k^{\frac{1}{2}}\right)\right)^2 S_k^2\right)^{-1}A_k^{\frac{1}{2}}\mathbf{x}\\
&\leq \left(A_k^{\frac{1}{2}}\mathbf{x}\right)^T\left(1-S_k^2\right)^{-1}A_k^{\frac{1}{2}}\mathbf{x}\\
&\leq (1-q^2)^{-1}\,\mathbf{x}^T A_k\mathbf{x},
\end{aligned}$$

which proves (6.4).

Let $\mathbf{x} \in U_2$ and $\mathbf{x} \in \text{Null}^{\perp}(A_k)$. Then, using similar arguments as above, using the fact that $\overline{M}_k^{-1}$ is s.p.d. on $\text{Null}^{\perp}(A_k)$, we have

$$\begin{aligned}
\mathbf{x}^T \overline{M}_k^{-1} \mathbf{x} &= \mathbf{x}^T A_k^{-1} \left[I - \left(I - \frac{1}{b_{k+1}} A_k S_k^2 \right)^2 S_k^2 \right] \mathbf{x} \\
&\geq \mathbf{x}^T A_k^{-1} \left(I - \left(I - \frac{1}{b_{k+1}} A_k S_k^2 \right) \right) \mathbf{x} \\
&= \frac{1}{b_{k+1}} \|S_k \mathbf{x}\|^2 \\
&\geq \frac{q^2}{b_{k+1}} \|\mathbf{x}\|^2.
\end{aligned}$$

Therefore,

$$\mathbf{x}^T \overline{M}_k^{-1} \mathbf{x} \geq \frac{q^2}{b_{k+1}} \|\mathbf{x}\|^2 \text{ for } \mathbf{x} \in U_2,$$

and (6.5) follows since U_2 is an invariant subspace of $\overline{M}_k$ as well.

To prove (6.6) for $\mathbf{x} \in \text{Null}^{\perp}(A_k)$, we write $\mathbf{x} \in \mathbb{R}^{n_k}$ as the ℓ_2– and $\overline{M}_k$–orthogonal sum

$$\mathbf{x} = \mathbf{x}_1 + \mathbf{x}_2, \quad \mathbf{x}_1 \in U_1,\ \mathbf{x}_2 \in U_2.$$

Then using (6.4), (6.5) and the obvious fact that $\|\mathbf{x}_1\|_{A_k} \leq \|\mathbf{x}\|_{A_k}$ and $\|\mathbf{x}_2\| \leq \|\mathbf{x}\|$, we have

$$\begin{aligned}
\|\mathbf{x}\|_{\overline{M}_k}^2 &= \|\mathbf{x}_1 + \mathbf{x}_2\|_{\overline{M}_k}^2 = \|\mathbf{x}_1\|_{\overline{M}_k}^2 + \|\mathbf{x}_2\|_{\overline{M}_k}^2 \\
&\leq \frac{1}{1-q^2} \|\mathbf{x}_1\|_{A_k}^2 + \frac{b_{k+1}}{q^2} \|\mathbf{x}_2\|^2 \\
&\leq \frac{1}{1-q^2} \|\mathbf{x}\|_{A_k}^2 + \frac{b_{k+1}}{q^2} \|\mathbf{x}\|^2,
\end{aligned}$$

completing the proof of (6.6).

A final observation is that if we extend $\overline{M}_k$ on $\mathbb{R}^{n_k}$ as semi–definite matrix vanishing on Null (A_k), then estimates (6.4), (6.5), and (6.6) still hold. □

Lemma 6.2. *Let s_ν denote the function φ_ν given by* (3.1) *with $b = 1$, i.e.*

$$s_\nu(t) = (-1)^\nu \frac{1}{2\nu+1} \frac{T_{2\nu+1}(\sqrt{t})}{\sqrt{t}}, \tag{6.7}$$

where T_n is the Chebyshev polynomial of degree n. Consider the quantity (estimated in Proposition 7.3 in the Appendix)

$$C_{1,\nu} \equiv \min_{t \in [0,1]} \left[1 - (2\nu+1)^2 s_\nu^2(t) t \right] s_\nu(t) \geq -1 + \delta_0 > -1 \quad \forall \nu, \tag{6.8}$$

where $\delta_0 \in (0, 1]$ is a constant independent of ν. Then, it holds that

$$\mathbf{x}^T M_k^{-1} A_k \left(2I - M_k^{-1} A_k \right)^{-1} \mathbf{x} \leq \frac{2 - \delta_0}{\delta_0} \|\mathbf{x}\|^2.$$

Finally, we show that M_k can be extended on $\mathbb{R}^{n_k}$ as semi–definite matrix that satisfies the estimate

$$\mathbf{x}^T \left(M_k + M_k^T - A_k\right) \mathbf{x} = \mathbf{x}^T \left(2M_k - A_k\right) \mathbf{x} \geq \frac{\delta_0}{2-\delta_0} \mathbf{x}^T A_k \mathbf{x}.$$

Equivalently, we have

$$\mathbf{x}^T A_k \left(M_k + M_k^T - A_k\right)^{-1} A_k \mathbf{x} \equiv \mathbf{x}^T A_k (2I - M_k^{-1} A_k)^{-1} M_k^{-1} A_k \mathbf{x} \leq \frac{2-\delta_0}{\delta_0} \mathbf{x}^T A_k \mathbf{x}.$$

That is, estimate (2.7) *holds with* $\alpha = \frac{\delta_0}{2-\delta_0}$.

Proof. First we observe that the prolongator smoother S_k defined by $S_k = \varphi_\nu(A_k)$ with $b = b_k$ satisfies

$$S_k = s_\nu(b_k^{-1} A_k).$$

Then for the error propagation operator (6.1) of our multigrid smoother, we have

$$\begin{aligned} I - M_k^{-1} A_k &= \left(I - \frac{1}{b_{k+1}} A_k S_k^2\right) S_k \\ &= \left(I - \frac{(2\nu+1)^2}{b_k} S_k^2 A_k\right) S_k \\ &= \left[I - (2\nu+1)^2 s_\nu^2 \left(\frac{1}{b_k} A_k\right) \left(\frac{1}{b_k} A_k\right)\right] s_\nu \left(\frac{1}{b_k} A_k\right). \end{aligned}$$

This shows that the product $M_k^{-1} A_k$ (which is well-defined without actually defining M_k) is a symmetric matrix that satisfies

$$\mathbf{x}^T \left(I - M_k^{-1} A_k\right) \mathbf{x} \geq \min_{t \in [0,1]} \left[1 - (2\nu+1)^2 s_\nu^2(t) t\right] s_\nu(t) \|\mathbf{x}\|^2.$$

Thus, by assumption of the Lemma, we have $\mathbf{x}^T (I - M_k^{-1} A_k) \mathbf{x} \geq (-1 + \delta_0) \|\mathbf{x}\|^2$ with $\delta_0 \in (0,\ 1]$. This implies that the symmetric matrix $2I - M_k^{-1} A_k$ is coercive, i.e.,

$$\mathbf{x}^T \left(2I - M_k^{-1} A_k\right) \mathbf{x} \geq \delta_0 \|\mathbf{x}\|^2.$$

Hence $2I - M_k^{-1} A_k$ is invertible. The expression of interest is

$$M_k^{-1} A_k \left(2I - M_k^{-1} A_k\right)^{-1} = -I + 2 \left(2I - M_k^{-1} A_k\right)^{-1}.$$

Since both matrices $M_k^{-1} A_k$ and $\left(2I - M_k^{-1} A_k\right)^{-1}$ are symmetric, non-negative, and commute, the product is also a non-negative matrix. We also have the estimate

$$\mathbf{x}^T M_k^{-1} A_k \left(2I - M_k^{-1} A_k\right)^{-1} \mathbf{x} \leq \left(-1 + \frac{2}{\delta_0}\right) \|\mathbf{x}\|^2 = \frac{2-\delta_0}{\delta_0} \|\mathbf{x}\|^2, \tag{6.9}$$

proving the first desired result. Based on the definition of $\overline{M}_k$ as symmetric semi-definite matrix with Null $(\overline{M}_k)$ = Null (A_k), we can define M_k on $\mathbb{R}^{n_k}$ using the expression

$$M_k = \overline{M}_k \left(2I - M_k^{-1} A_k\right),$$

which shows that M_k is invertible on Null$^\perp$ (A_k) as a product of two invertible matrices. It also shows that M_k is positive definite on Null$^\perp$ (A_k) since the two terms (with common set of eigenvectors) are positive definite (on Null$^\perp$ (A_k)). The final result follows from (6.9), first for vectors in Null$^\perp$ (A_k), using the fact that M_k is s.p.d. and the

ν	$C_\nu/(2\nu+1)$	$C_{1,\nu}$	ν	$C_\nu/(2\nu+1)$	$C_{1,\nu}$
1	0.444444444444	-0.135229953887	120	0.318318901912	-0.103057437637
2	0.344265186330	-0.113196279492	130	0.318317573036	-0.103056836776
3	0.330166016890	-0.108055102603	140	0.318316517698	-0.103056359520
4	0.325185126688	-0.106037861689	150	0.318315665665	-0.103055974286
5	0.322817680242	-0.105037539348	160	0.318314967883	-0.103055658768
6	0.321499599981	-0.104468396378	170	0.318314389245	-0.103055396432
7	0.320688247098	-0.104113591413	180	0.318313904096	-0.103055175377
8	0.320152568605	-0.103877442695	190	0.318313493321	-0.103054991898
9	0.319780050865	-0.103712323026	200	0.318313142470	-0.103054831135
10	0.319510389628	-0.103592331317	210	0.318312840427	-0.103054696141
20	0.318622268380	-0.103194362894	220	0.318312578538	-0.103054576126
30	0.318450784453	-0.103117025348	230	0.318312349991	-0.103054474895
40	0.318389749971	-0.103117025348	240	0.318312149354	-0.103054382088
50	0.318361238793	-0.103076576949	250	0.318311972264	-0.103054295161
60	0.318345660541	-0.103069535619	260	0.318311815174	-0.103054232702
70	0.318336229255	-0.103065272070	270	0.318311675174	-0.103054164217
80	0.318330089757	-0.103062496345	280	0.318311549901	-0.103054100291
90	0.318325870932	-0.103060588868	290	0.318311437328	-0.103054061698
100	0.318322847786	-0.103059221879	300	0.318311335785	-0.103054013059
110	0.318320607725	-0.103058208892	–	–	–

TABLE 1. Numerical evaluation of the constants $\frac{C_\nu}{2\nu+1}$ and $C_{1,\,\nu}$

commutativity of all terms. Finally, for vectors in Null (A_k), the result is trivially seen based on the definition of M_k. □

We summarize numerical evaluation of the constants $C_{1,\,\nu}$ defined in (6.8), and $\frac{C_\nu}{2\nu+1}$ defined in Proposition 3.3, in Table 1.

7. Numerical experiments and remarks on complexity and aggregation strategies

Myriad algebraic aggregation schemes have been proposed in the literature (e.g., [SA4th, SA]). For isotropic problems, these schemes typically result in a method with very low storage and computational complexities, and the coarse-level operator fill-in typically remains nicely bounded. However, proving that this is so is, in general, difficult. Indeed, although overall well behaved, the aggregates obtained by these schemes may contain, depending on the numbering of unknowns, aggregates that are not ideally aligned even for uniform meshes.

We comment next on one way of constructing aggregates that leads to coarse matrices with controlled sparsity pattern. Namely, assume we are given a quasi-uniform mesh $\mathcal{T}_h$ that triangulates our polygonal (or polyhedral) domain Ω. Choose a parameter H and generate a uniform mesh $\mathcal{T}_H$ with boxes of size $H \times H$ ($\times H$ in 3D). Consider only those boxes that provide covering of Ω. Each box Ω_{ij}(or Ω_{ijk} in 3D) intersects part of the mesh

$\mathcal{T}_h$. In this way we construct aggregates $\mathcal{A}_{ij}$ (or $\mathcal{A}_{ijk}$) each containing all fine–grid vertices that are within a particular box (with some arbitration of nodes on box boundaries if any). The only requirement is that the resulting aggregates have large enough interior which can be ensured if H is large enough and $\mathcal{T}_h$ is fine enough. Then, if we use one step of SA with the thus constructed aggregates, the resulting coarse matrix A_H will have, in 2D, the sparsity pattern of a 9-point stencil (27-point in 3D). The situation is illustrated in Figures 1- 2. In general, this type of coarsening may result in fairly large coarsening factor H/h. To compensate for it, we can use the polynomial smoother that we analyzed in the preceding section.

From that level on, we are essentially working with a finite difference matrix on a regular mesh with lexicographic numbering, so the complexity cost of the resulting SA procedure is straightforward to handle; it is essentially the same as of a geometric multigrid on uniformly refined mesh.

Finally, we comment that this type of coarsening ensures uniform constant σ_a in our main assumption (4.3) (or uniform Poincaré constant c_a in (4.1)).

In what follows, we present several performance tests of the SA method based on the aggregation procedure explained above. At the initial level, we use aggressive coarsening. Also, at that level, we use higher order polynomial degree, ν_0^P, for the smoothed prolongator, and ν_0^R for the relaxation method. We use three meshes (Mesh6, Mesh7, and Mesh8) obtained by successive steps of uniform refinement of the unstructured mesh shown in Fig. 1. We also show a 3D example that corresponds to a deformed cubic domain. The respective results are shown in Tables 2–4 for the 2D examples, and in Table 5 for the 3D one. The last two columns of the tables show the relative condition number of the SA method with respect to the given matrix and the operator complexity (a standard AMG measure defined as the sum of the nonzero entries of the matrices at all coarsening levels divided by the number of nonzero entries of the fine-grid matrix).

For efficiency reason, the polynomial used in the relaxation process at the initial level was implemented in the following split form. Based on the fact that the relaxation iteration matrix has the product form $I - M^{-1}A = (I - M_1^{-1}A)(I - M_2^{-1}A) = s_\nu(b^{-1}A)\left[I - T_{2\nu+1}^2\left(\sqrt{b^{-1}A}\right)\right]$ $(b \geq \|A\|)$, we used $s_\nu(t)$ in the pre–relaxation and $1 - T_{2\nu+1}^2(\sqrt{t})$ in the post-relaxation process. This is only at the initial level. In general, the method can be implemented (using proper restriction matrix $R \neq P^T$) such that the overall V-cycle defines a s.p.d. SA. In our tests, the V-cycle leads to a nonsymmetric operator, B^{-1}, that is coercive and provides an accurate approximate inverse to A. We ran the method as a preconditioner in a Krylov type method (which in the s.p.d. case reduces to the conjugate gradient method). The condition number estimates that we provide are simply approximations to $1/(1-\varrho)$ where ϱ is the convergence factor.

In all tables below we used the following stopping criterion:

$$\frac{\mathbf{z}^T\mathbf{r}_k}{\mathbf{z}_0^T\mathbf{r}_0} \leq \varepsilon\sqrt{\kappa}, \tag{7.1}$$

where $\mathbf{r}_k$ denotes the residual at iteration k, $\mathbf{r}_k = \mathbf{f} - A\mathbf{x}_k$, $\mathbf{z}_k = B^{-1}\mathbf{r}_k$ denotes the preconditioned residual at iteration k, κ denotes the condition number of the preconditioned system $B^{-1}A$, and $\varepsilon = 10^{-6}$. Some of the tables list two iteration counts. In such cases,

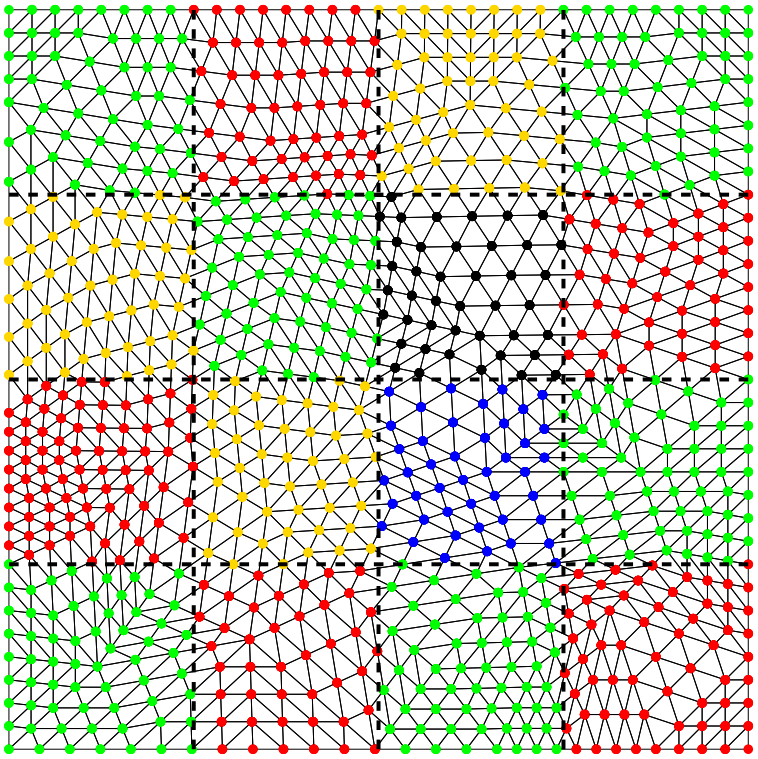

FIGURE 1. Formation of aggregates to guarantee sparsity of all coarse-level operators.

the first number is the number of iterations performed to reach relative residual smaller than ε, the second is the number of iterations to satisfy (7.1). The reported timings correspond to the higher iteration count. All experiments were carried out on a notebook personal computer with 2 GHz Intel Core2 Duo P7350 CPU and 4 GB of RAM.

All tables clearly demonstrate that the SA with aggressive coarsening performs as expected; it has bounded condition number and respective number of iterations when we vary the size of the fine-grid problem. The setup and solution timings for SA with aggressive coarsening are higher than SA with more gradual coarsening. The geometric way of coarsening is overall less efficient (in terms of timings) than the well–tuned matrix-based

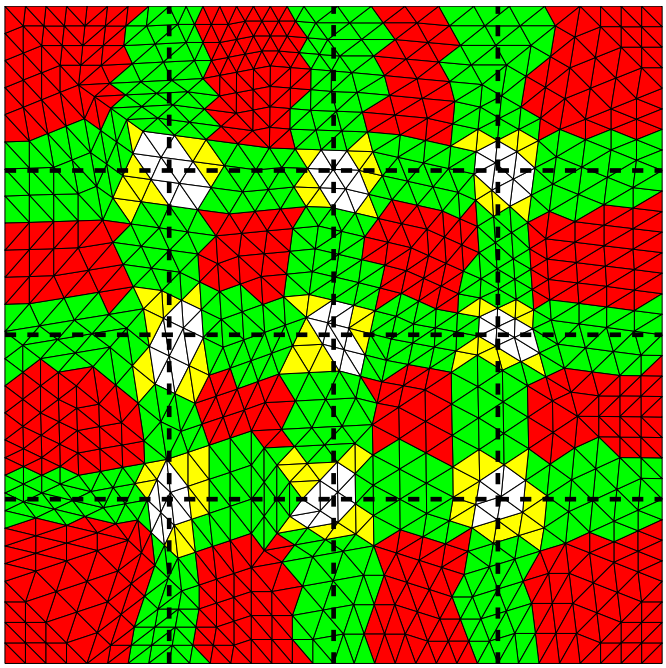

FIGURE 2. The overlap of the extended aggregates obtained by applying two actions of A illustrating the sparsity of the resulting SA coarse-level operator. Darker color corresponds to elements that intersect fewer extended aggregates.

1^{st} coarse level	# levels	ν_0^P	ν_0^R	Setup time	Iteration time	Iterations	Cond	Oper. Cmplx.
64	3	12	12	0.463	0.823	8	2.467	1.00205
64	3	12	11	0.463	0.845	8/9	2.675	1.00205
64	3	12	10	0.463	0.777	9	2.958	1.00205
625	4	5	5	0.315	0.411	8	2.453	1.03174
625	4	5	4	0.312	0.376	9	2.513	1.03174
625	4	5	3	0.309	0.373	10/11	3.497	1.03174
625	4	4	4	0.266	0.373	9	2.608	1.02555
625	4	4	3	0.247	0.373	10/11	3.566	1.02555
5625	4	3	3	0.578	0.302	7	1.867	1.54096
5625	4	3	2	0.557	0.283	8	1.894	1.54096
5625	4	3	1	0.558	0.250	9	2.338	1.54096
5625	4	2	2	0.337	0.288	8/9	2.477	1.36543
5625	4	2	1	0.337	0.225	9	2.405	1.36543
5625	4	1	1	0.272	0.214	9/10	2.655	1.19211
90434	4	1	1	0.265	0.241	9	3.003	1.17419
90434	4	1	1	0.271	0.263	12/13	5.533	1.17419

TABLE 2. Results for 2D unstructured problem with $51,681$ degrees of freedom (Mesh6); Dirichlet BC were imposed at 964 of its boundary nodes. The last 2 lines correspond to standard SA solver with default aggregation and Gauss-Seidel and Jacobi relaxation, respectively, used on all levels. For $k > 0$, values $\nu_k = 1$ have been used.

1st coarse level	# levels	ν_0^P	ν_0^R	Setup time	Iteration time	Iterations	Cond	Oper. Cmplx.
289	3	12	12	2.220	4.645	8/9	2.693	1.00282
289	3	12	11	2.232	4.287	9	2.917	1.00282
289	3	12	10	2.218	4.331	9/10	3.337	1.00282
289	3	9	11	1.597	4.658	9/10	3.578	1.00199
289	3	9	10	1.591	4.631	10/11	3.834	1.00199
289	3	9	9	1.607	4.233	11	4.276	1.00199
289	3	9	8	1.594	4.174	11/12	4.972	1.00199
2500	4	6	6	1.757	2.353	8	2.530	1.03879
2500	4	6	5	1.754	2.313	8/9	2.625	1.03879
2500	4	6	4	1.759	1.919	8/9	2.554	1.03879
2500	4	6	3	1.732	1.902	10/11	3.643	1.03879
22500	5	3	3	2.386	1.922	8/9	2.682	1.55218
22500	5	3	2	2.345	1.557	9	2.713	1.55218
22500	5	3	1	2.347	1.228	9	2.545	1.55218
22500	5	2	2	1.457	1.643	10	3.457	1.37477
22500	5	2	1	1.453	1.249	10	3.174	1.37477
22500	5	1	1	1.220	1.100	10	2.906	1.19465
17271	5	1	1	1.020	1.310	9/10	3.692	1.10914
17271	5	1	1	1.022	1.243	13/14	5.741	1.10914

TABLE 3. Results for 2D unstructured problem with $205,761$ degrees of freedom (Mesh7); Dirichlet BC were imposed at all $1,924$ boundary nodes. The last 2 lines correspond to standard SA solver with default aggregation and Gauss-Seidel and Jacobi relaxation, respectively, used on all levels. For $k > 0$, values $\nu_k = 1$ have been used.

1st coarse level	# levels	ν_0^P	ν_0^R	Setup time	Iteration time	Iterations	Cond	Oper. Cmplx.
144	3	30	30	17.876	46.878	8/9	3.210	1.00028
144	3	30	25	17.669	42.745	9/10	4.059	1.00028
144	3	30	20	17.669	41.570	11/12	5.582	1.00028
1156	4	13	13	9.728	21.462	8/9	2.889	1.00324
1156	4	13	12	9.743	20.038	8/9	2.891	1.00324
1156	4	13	11	9.700	18.432	9	3.035	1.00324
1156	4	13	10	9.647	18.460	9/10	3.527	1.00324
10201	4	6	6	7.178	11.344	9	3.048	1.04092
10201	4	6	5	7.121	9.781	9	3.023	1.04092
10201	4	6	4	7.087	8.284	9	3.020	1.04092
10201	4	6	3	7.049	8.135	10/11	3.652	1.04092
10201	4	6	2	7.132	8.621	14/15	6.757	1.04092
10201	5	2	2	5.930	7.741	10/11	3.900	1.38440
10201	5	2	1	5.930	5.899	10/11	3.571	1.38440
10201	5	1	1	4.798	7.741	10	3.072	1.19946
90434	5	1	1	4.891	6.655	10/11	5.076	1.1787
90434	5	1	1	4.950	6.669	14/15	9.007	1.1787

TABLE 4. Results for 2D unstructured problem with $821,121$ degrees of freedom (Mesh8); Dirichlet BC were imposed at $3,844$ of its boundary nodes. The last 2 lines correspond to standard SA solver with default aggregation and Gauss-Seidel and Jacobi relaxation, respectively, used on all levels. For $k > 0$, values $\nu_k = 1$ have been used.

aggregation algorithms developed previously in earlier works, but still shows reasonably competitive performance (for small polynomial degrees ν_0^R and ν_0^P). Finally, we note that the polynomial smoothers and the (scaled) Jacobi are the ones that are straightforward to parallelize and hence of more practical interest. At the end, we comment that there is some room to optimize the SA setup in the case of large polynomial degrees that needs further study.

1st coarse level	# levels	ν_0^P	ν_0^R	Setup time	Iteration time	Iterations	Cond	Oper. Cmplx.
64	3	8	8	30.174	30.799	7	2.118	1.00003
64	3	8	7	30.132	31.044	8	2.453	1.00003
64	3	8	6	30.059	30.131	9	3.045	1.00003
64	3	8	5	30.047	31.231	10/11	4.074	1.00003
64	3	6	6	23.425	29.974	9	3.192	1.00003
64	3	6	5	23.359	31.061	10/11	4.117	1.00003
1680	4	4	4	27.344	17.737	7	2.266	1.00159
1680	4	4	3	27.339	14.569	7	1.927	1.00159
1680	4	4	2	27.499	14.178	9	2.557	1.00159
1680	4	4	1	27.441	15.475	13/14	6.172	1.00159
1680	4	3	3	20.595	18.053	8/9	2.673	1.00111
1680	4	3	2	20.620	14.005	9	2.644	1.00111
1680	4	3	1	20.585	14.978	13/14	6.159	1.00111
46248	4	2	2	44.717	11.461	6	1.560	1.14223
46248	4	2	1	44.426	8.494	6	1.492	1.14223
46248	4	1	1	22.701	8.551	7	1.935	1.04375
51266	4	1	1	27.214	16.372	7	2.394	1.10982
51266	4	1	1	27.106	11.467	9/10	3.449	1.10982

TABLE 5. Results for 3D problem with $1,367,631$ degrees of freedom; Dirichlet BC were imposed only on one of the boundary faces. The last 2 lines correspond to standard SA solver with default aggregation and Gauss-Seidel and Jacobi relaxation, respectively, used on all levels. For $k > 0$, values $\nu_k = 1$ have been used.

APPENDIX

Here, we study the behavior of the quantities $\frac{C_\nu}{2\nu+1}$ and $C_{1,\nu}$ defined in (3.5) and (6.8), respectively, for large values of ν. For this reason, we consider the remainders Q_k and P_k of the Chebyshev polynomials ($T_0 = 1,\ T_1(t) = t$ and $T_{i+1} = 2tT_i - T_{i-1}$, for $i \geq 1$), expanded as follows

$$\begin{aligned} T_{2k-1}(t) &= (-1)^{k-1}(2k-1)t + tQ_{k-1}(t^2),\ Q_0 = 0, \\ T_{2k}(t) &= (-1)^k + P_k(t^2),\ P_0 = 0,\ P_1(t^2) = 2t^2. \end{aligned} \tag{7.2}$$

Using induction (cf. Proposition 6.25 in [Va08]), the following recursive relations hold.

Proposition 7.1.

$$\begin{aligned} Q_k(t^2) &= (4t^2-1)Q_{k-1}(t^2) - 2P_{k-1}(t^2) + 4(2k-1)(-1)^{k-1}t^2, \\ P_k(t^2) &= -P_{k-1}(t^2) + 2(-1)^{k-1}(2k-1)t^2 + 2t^2Q_{k-1}(t^2). \end{aligned} \tag{7.3}$$

Proof. Indeed, assuming that (7.2) hold for some $k \geq 1$. Then, using the formula $T_{2k+1} = 2tT_{2k} - T_{2k-1}$, we have

$$T_{2k+1} = 2t((-1)^k + P_k(t^2)) - (-1)^{k-1}(2k-1)t - tQ_{k-1}(t^2) = (-1)^k(2k+1)t + t\left(2P_k(t^2) - Q_{k-1}(t^2)\right).$$

Thus, $Q_k(t^2) = 2P_k(t^2) - Q_{k-1}(t^2)$. Similarly, we have

$$\begin{aligned} T_{2k+2} &= 2tT_{2k+1} - T_{2k} \\ &= 2t\left((-1)^k(2k+1)t + tQ_k(t^2)\right) - (-1)^k - P_k(t^2) \\ &= (-1)^{k+1} + \left((-1)^k(2k+1)2t^2 + 2t^2Q_k(t^2) + P_k(t^2)\right). \end{aligned}$$

That is, we can set $P_{k+1}(t^2) = -P_k(t^2) + (-1)^k(2k+1)2t^2 + 2t^2Q_k(t^2)$. Finally from $Q_k(t^2) = 2P_k(t^2) - Q_{k-1}(t^2) = 2\left(-P_{k-1}(t^2) + (-1)^{k-1}(2k-1)2t^2 + 2t^2Q_{k-1}(t^2)\right) - Q_{k-1}(t^2)$,

we obtain

$$Q_k(t^2) = -2P_{k-1}(t^2) + (-1)^{k-1}(2k-1)4t^2 + (4t^2-1)Q_{k-1}(t^2).$$

□

Our first main result is the following estimate.

Proposition 7.2. *The quantity*

$$C_\nu = \sup_{t\in(0,\,1]} \frac{\left|1-(-1)^\nu \frac{T_{2\nu+1}(\sqrt{t})}{(2\nu+1)\sqrt{t}}\right|}{\sqrt{t}} = \sup_{t\in(0,\,1]} \frac{|(-1)^\nu(2\nu+1)t - T_{2\nu+1}(t)|}{(2\nu+1)t^2},$$

satisfies the bound

$$\frac{C_\nu}{2\nu+1} \le 2.$$

Proof. This is also the proof of Proposition 3.4.

We notice first that, if t is away from the origin, i.e., $t(2\nu+1) \ge 1$, we have

$$\sup_{t\in(\frac{1}{2\nu+1},\,1]} \frac{|(-1)^\nu(2\nu+1)t - T_{2\nu+1}(t)|}{t^2(2\nu+1)^2} \le 1 + \sup_{t\in(0,\,1]} |T_{2\nu+1}(t)| \le 2. \tag{7.4}$$

The quantity of interest, $\frac{C_\nu}{2\nu+1}$ can be expressed in terms of the remainder Q_ν as follows

$$\frac{C_\nu}{2\nu+1} = \sup_{t\in(0,1]} \frac{|(-1)^\nu(2\nu+1)t - T_{2\nu+1}(t)|}{t^2(2\nu+1)^2} = \sup_{t\in(0,1]} \frac{|Q_\nu(t^2)|}{t(2\nu+1)^2}.$$

We continue with the estimation of the last quantity. Our next goal is to show that for

$$\frac{|Q_k(t^2)|}{t(2k+1)} \le 2k \text{ for any } k \le \nu \text{ and } t \in (0,\ \frac{1}{2\nu+1}]. \tag{7.5}$$

This, together with (7.4), guarantees the desired uniform bound on $\frac{C_\nu}{2\nu+1}$.

We assume that $\nu \ge 1$. This implies that $1-4t^2 > 0$, i.e., $4t^2 \le \frac{4}{(2\nu+1)^2} < 1$.

Introduce the expression

$$R_k = 2P_k - (-1)^k(2k+1)4t^2.$$

The recursion (7.3) for P_k can be rewritten in terms of R_k and Q_k. We have

$$\begin{aligned} R_k &= -R_{k-1} + 4t^2Q_{k-1} - (-1)^k(2k+1)4t^2 \\ &= -R_{k-1} + 4t^2Q_{k-1} + (-1)^{k-1}(2k-1)4t^2 + 8(-1)^k t^2 \\ &= -R_{k-1} + 4\left(t^2Q_{k-1} + (-1)^{k-1}(2k-1)t^2\right) + 8(-1)^k t^2 \\ &= -R_{k-1} + 4tT_{2k-1} + 8(-1)^k t^2. \end{aligned}$$

Therefore, the following estimate holds

$$\frac{|R_k|}{t} \le \frac{|R_{k-1}|}{t} + 4|T_{2k-1}| + 8t \le \frac{|R_{k-1}|}{t} + 4t(2k-1) + 8t \le \frac{|R_{k-1}|}{t} + 4,$$

and since $R_0 = 0$, we obtain

$$|R_k| \le 4kt.$$

The desired estimate for Q_k follows from the representation

$$Q_k = -R_k + (4t^2-1)Q_{k-1}.$$

We have

$$\frac{|Q_k|}{t(2k+1)} \le \frac{(1-4t^2)(2k-1)}{2k+1}\,\frac{|Q_{k-1}|}{t(2k-1)} + \frac{|R_k|}{t(2k+1)} \le \frac{|Q_{k-1}|}{t(2k-1)} + \frac{4k}{2k+1} \le \frac{|Q_{k-1}|}{t(2k-1)} + 2.$$

That is, since $Q_0 = 0$, we obtain

$$\frac{|Q_k|}{t(2k+1)} \le 2k < 2\nu + 1.$$

Combining the two estimates, (7.4) for $t > \frac{1}{2\nu+1}$ and (7.5) for $t \le \frac{1}{2\nu+1}$, we obtain the uniform bound

$$\frac{C_\nu}{2\nu+1} \le 2.$$

□

Numerical experiments presented in Table 1, show that for large ν the quantity behaves asymptotically as $\frac{C_\nu}{2\nu+1} \simeq 0.318311...$

Next, we study the quantity $C_{1,\,\nu}$ defined in (6.8).

The polynomial smoother (6.1) leads to the polynomial

$$\left[1 - (2\nu+1)^2 s_\nu^2(t) t\right] s_\nu(t),$$

where

$$s_\nu(t) = (-1)^\nu \frac{1}{2\nu+1} \frac{T_{2\nu+1}(\sqrt{t})}{\sqrt{t}},$$

This shows that

$$1 - (2\nu+1)^2 s_\nu^2(t) t = 1 - T_{2\nu+1}^2(\sqrt{t}).$$

Since both $\sqrt{t}$ and t vary in $[0,1]$, we can substitute t for $\sqrt{t}$ and in what follows we estimate the expression

$$1 + \left(1 - T_{2k+1}^2(t)\right)(-1)^k \frac{T_{2k+1}(t)}{t(2k+1)}.$$

Our goal is to prove a positive lower bound $\delta_0 \in (0,1]$.

For $t \in [\frac{1}{2k+1},\ 1]$, use the inequality

$$\begin{aligned} 1 + \left(1 - T_{2k+1}^2(t)\right)(-1)^k \tfrac{T_{2k+1}(t)}{t(2k+1)} &\ge 1 - \left(1 - T_{2k+1}^2(t)\right) \tfrac{|T_{2k+1}(t)|}{t(2k+1)} \\ &\ge 1 - \left(1 - T_{2k+1}^2(t)\right) |T_{2k+1}(t)| \\ &\ge 1 - \max_{x\in[0,1]}\ x(1-x^2) \\ &= 1 - \tfrac{2}{3\sqrt{3}} > 0. \end{aligned} \tag{7.6}$$

For $t \in (0,\ \frac{1}{2k+1}]$ use the representation

$$(-1)^k \frac{T_{2k+1}(t)}{t(2k+1)} = 1 + (-1)^k \frac{Q_k(t^2)}{2k+1}.$$

Based on estimate (7.5), we then have for $t \in (0,\ \frac{1}{2k+1}]$,

$$\frac{|Q_k(t^2)|}{2k+1} \le 2kt \le \frac{2k}{2k+1},$$

which implies

$$\left(1-T_{2k+1}^2(t)\right)(-1)^k\frac{T_{2k+1}(t)}{t(2k+1)}=\left(1-T_{2k+1}^2(t)\right)\left(1+(-1)^k\frac{Q_k(t^2)}{2k+1}\right)\geq\frac{1-T_{2k+1}^2(t)}{2k+1}\geq 0.$$

Therefore for $t\in(0,\ \frac{1}{2k+1}]$, we have

$$1+\left(1-T_{2k+1}^2(t)\right)(-1)^k\frac{T_{2k+1}(t)}{t(2k+1)}\geq 1. \tag{7.7}$$

Combining estimates (7.6) and (7.7), we obtain the following result, which is a verification of Proposition 3.5.

Proposition 7.3. *The following coercivity estimate holds:*

$$1+\left(1-T_{2k+1}^2(t)\right)(-1)^k\frac{T_{2k+1}(t)}{t(2k+1)}\geq\delta_0,$$

where $\delta_0=1-\frac{2}{3\sqrt{3}}\in(0,1)$.

By numerical experiments (see Table 1), a better lower bound of 0.89 for $\delta_0=\min\limits_{\nu\geq 1}(1+C_{1,\,\nu})$ can be obtained.

References

[BD96] F. A. Bornemann and P. Deuflhard, "*The cascadic multigrid method for elliptic problems,*" Numerische Mathematik **75**(1996), pp. 135-152.

[BHMV] M. Brezina, C. Heberton, J. Mandel, and P. Vaněk, "*An iterative method with convergence rate chosen a priori,*" UCD CCM Report **140**(1999), available at http://www-math.cudenver.edu/ jmandel/papers/.

[TD06] T. A. Davis, "*Direct Methods for Sparse Linear Systems,*" SIAM, Philadelphia, 2006. 217 p.

[Sha94] V. V. Shaidurov, "*Multigrid Methods for Finite Elements,*" Kluwer Academic Publishers, Dordrecht, 1995. p. 331.

[KrVa] J. Křížková, P. Vaněk, *"Two–level preconditioner with small coarse grid appropriate for unstructured meshes,"* Numerical Linear Algebra with Applications **3**(1996), no. 4, pp. 255–274.

[VanSA] P. Vaněk, "*Acceleration of convergence of a two-level algorithm by smoothing transfer operator,*" Applications of Mathematics **37**(1992), pp. 265–274.

[Van] P. Vaněk *"Smoothed prolongation multigrid with rapid coarsening and massive smoothing",* Applications of Mathematics (to appear).

[SA] P. Vaněk, M. Brezina, and J. Mandel, "*Convergence of algebraic multigrid based on smoothed aggregation,*" Numerische Mathematik **88**(2001), pp. 559–579.

[VBT] P. Vaněk, M. Brezina, and R. Tezaur, *"Two-grid method for elasticity on unstructured meshes,"* SIAM Journal on Scientific Computing **21**(1999), no. 3, pp. 900–923.

[SA4th] P. Vaněk, J. Mandel, and M. Brezina, "*Algebraic multigrid by smoothed aggregation for second and fourth order elliptic problems,*" Computing **56**(1996), pp. 179–196.

[Va08] P. S. Vassilevski, "*Multilevel Block Factorization Preconditioners,* Matrix-based Analysis and Algorithms for Solving Finite Element Equations," Springer, New York, 2008. 514 p.

[XZ] J. Xu and L. T. Zikatanov, "*The method of alternating projections and the method of subspace corrections in Hilbert space,*" Journal of the American Mathematics Society **15**(2002), pp. 573-597.

Department of Applied Mathematics, Campus Box 526, University of Colorado at Boulder, Boulder, CO 80309-0526, U.S.A.
E-mail address: marian.brezina@gmail.com

E-mail address: petrvanek09@seznam.cz

Center for Applied Scientific Computing, Lawrence Livermore National Laboratory, P.O. Box 808, L-560, Livermore, CA 94550, U.S.A.
E-mail address: panayot@llnl.gov